



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
28.11.2001 Bulletin 2001/48

(51) Int Cl.7: **C12N 15/12**, C12N 15/62,
C07K 14/47, C07K 16/18,
C12Q 1/68, G01N 33/53,
A61K 38/17, A61K 31/7088

(21) Application number: **01112637.2**

(22) Date of filing: **24.05.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

- **Ji, Yonggang**
San Mateo, CA 94403 (US)
- **Penn, Sharron G.**
San Mateo, CA 94402 (US)
- **Hanzel, David K.**
Palo Alto, CA 94303 (US)

(30) Priority: **26.05.2000 US 207456 P**
05.02.2001 US 266860 P

(74) Representative: **Canning, Lewis R. et al**
Nycomed Amersham plc,
Amersham Laboratories,
White Lion Road
Amersham, Bucks HP7 9LL (GB)

(83) **Declaration under Rule 28(4) EPC (expert solution)**

(71) Applicant: **Aeomica, INC.**
Sunnyvale, CA 94086 (US)

Remarks:

The sequence listing, which is published as annex to the application documents, was filed after the date of filing. The applicant has declared that it does not include matter which goes beyond the content of the application as filed.

(72) Inventors:
• **Gu, Yizhong**
Sunnyvale, CA 94087 (US)

(54) **Myosin-like gene expressed in human heart and muscle**

(57) The invention provides compounds and compositions comprising isolated nucleic acids encoding a novel human myosin-like protein particularly expressed

in heart and muscle, compositions derivable directly or indirectly therefrom, including probes, proteins, and antibodies, and methods for using the same.

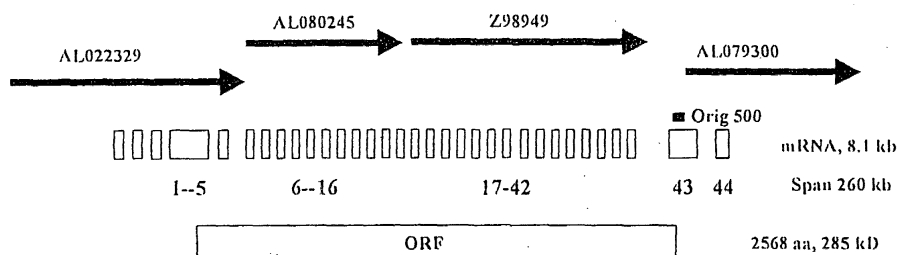


Fig. 5

Description

FIELD OF THE INVENTION

[0001] The present invention relates to isolated nucleic acids encoding a novel human myosin-like gene particularly expressed in heart and muscle, compounds and compositions derivable directly or indirectly therefrom, and methods for using the same.

BACKGROUND OF THE INVENTION

[0002] Myosins are proteins that act as ubiquitous intracellular engines, typically motoring along tracks of actin filaments to drive a variety of cellular processes including muscular contraction, cytokinesis, membrane trafficking and signal transduction. Baker *et al.*, *Curr. Opin. Cell Biol.* 10: 80-86 (1998). Given the range of intracellular passengers and itineraries, some form of the myosin protein is found in virtually all eukaryotic cells.

[0003] The myosin gene superfamily has at least seventeen members, or classes, encoded by multiple genes.

[0004] Mammalian cells have the largest number of myosin genes, with the 28 identified myosin genes belonging to nine classes. Sellers, *Biochim. Biophys. Acta* 1496: 3-22 (2000). The genome of the yeast, *Saccharomyces cerevisiae*, contains just 5 myosin genes. Brown, *Curr. Opin. Cell Biol.* 9: 44-48 (1997). Between these extremes, the nematode *Caenorhabditis elegans* has 14 identified myosin genes. Baker *et al.*, *J. Mol. Biol.* 172: 523-535 (1997). Myosins are also found in plant cells, with myosin genes in classes VIII, XI and XIII expressed exclusively in plants.

[0005] The structure of myosin always includes one or two heavy chains and several light chains.

[0006] The heavy chain is composed of three structurally and functionally different domains. The head domain, which is highly conserved across the myosin family, functions to generate force, and contains actin-binding and ATP-binding sites; the ATPase activity of the head domain is activated by actin binding. The neck region, which mediates association with the light chains, regulates the adjacent head domain. The tail domain regulates the specific activity of each myosin.

[0007] The light chains of myosin I and myosin V are calmodulin, which is a calcium-binding regulatory subunit in certain enzymes. Myosin II is also regulated by calcium-binding light chains, but not calmodulin.

[0008] Together, the disparate myosin heavy and light chains permit myosins to serve disparate cellular roles.

[0009] For example, in brush border microvilli, myosin I, one of the two most abundant forms of myosin, links the microfilament bundles to the plasma membrane. Myosin II, the other of the two most abundant myosin classes, forms dimers in muscle cells that associate to form thick filaments, which are part of the contractile apparatus.

[0010] In addition to its well characterized role in contraction and force production in skeletal, cardiac, and smooth muscles, myosin II is required for cytokinesis, cell motility, cell polarity/chemotaxis, maintaining cell architecture and development in nonmuscle cells. Sellers, *Biochim. Biophys. Acta* 1496:3-22 (2000).

[0011] Thus, contractile bundles, which comprise both actin and myosin II filaments, are found in numerous cell types. In epithelial cells, these bundles are called the circumferential belt and can function structurally (e.g., as an internal brace helping to control cell shape), or for motility (e.g., in wound healing, contraction seals the gap in a sheet of cells).

[0012] Myosin II is an integral component to the cytoskeletal substructure, acting to stiffen cortical membranes. In cytokinesis, myosin II plays an essential role performing the motor function for the contractile ring which constricts to form the cleavage furrow.

[0013] Myosin IXs, identified in rat, human and *C. elegans*, are expressed in a wide variety of tissues and cell types and are believed to be involved in intracellular signaling pathways. Myosin IX acts as a negative regulator of Rho (a G-protein), suggesting that it may control the Rho signaling pathways involved in cytoskeleton reorganization and other cellular processes. Bahler *et al.*, *Biochim. Biophys. Acta* 1496:52-59 (2000). However the precise cellular functions of the myosin IXs and their exact roles in the Rho signaling pathways have still to be determined.

[0014] Membrane-bound myosins of various classes, particularly myosin I and myosin V, have been implicated in movement of vesicles within the cell. Due to its co-localization with Golgi membrane, myosin I is thought to move membrane vesicles between membrane compartments in the cytoplasm. Myosin I also serves as a membrane-microfilament linkage in microvilli.

[0015] In yet other examples of myosin function, protein secretion in yeast is disrupted by mutation in the myosin V gene, suggesting an important role for myosin V. In vertebrate brain tissue, myosin V is concentrated in the Golgi stacks and at the tips of membrane processes extending from neuronal cells. Espreafico *et al.*, *J. Cell Biol.* 119: 1541 (1992). This type of membrane association would be consistent with the effects of myosin V mutations in mice, which adversely affect synaptic transmission, resulting in seizures and eventually death. In cell migration, different myosins localize to different regions of the cell. Myosin I localizes to the leading edge of a crawling amoeba, possibly participating in the translocation phase, whereas myosin II is more concentrated at the tail, where it is involved in retraction of the cell body.

[0016] Given their ubiquity, and the wide variety of tasks driven by myosins, it is not surprising that myosin defects

have been implicated in a wide variety of diseases.

[0017] For example, as noted above, myosin V mutations in mice adversely affect synaptic transmission, resulting in seizures and eventually death.

[0018] Myosin and myosin-like genes have also been implicated in a number of human diseases.

[0019] For example, myosin plays a role in hypertrophic cardiomyopathy. An autosomally dominant form of the disease is frequently caused by a missense point mutation in exon 13 of the cardiac myosin heavy chain gene on chromosome 14. Less often, an abnormal cardiac myosin heavy chain hybrid gene is present.

[0020] As another example, mutation in the gene encoding myosin VIIA is responsible for some forms of Usher syndrome. Weil *et al.*, *Nature* 374:60-61 (1995). Usher Syndrome is characterized by hearing impairment associated with retinitis pigmentosa.

[0021] Three classes of myosins, VI, VII and XV have been associated with genetic deafness disorders in mammals. Hasson, *Am. J. Hum. Genet.* 61:801-5 (1997); Redowicz, *J. Muscle Res. Cell Mot.* 20:241-248 (1999). Mutations in these myosins result in abnormalities in the stereocilia in the sensory cells of the inner ear of mice.

[0022] Other myosin-like genes may also be needed for normal stereociliary function in the inner ear, and mutations in these additional myosin-like genes may thus underlie or contribute to various forms of congenital deafness. For example, nonsyndromic hereditary deafness (DFNA17), mapped to chromosome 22q12.2-q13.3 in 1997, has now been associated with mutation in MYH9, a nonmuscle-myosin heavy-chain gene, located within the linked region. Lalwani *et al.*, *Am. J. Hum. Genet.* 67:1121-8 (2000).

[0023] Despite the long-standing interest in myosins and sequence similarity among classes and across species, not all myosin and myosin-like genes have yet been identified, even in species that are genetically well characterized.

[0024] For example, a novel myosin-like gene (MysPDZ) has recently been cloned from mouse bone marrow stromal cells. The protein encoded by this newly identified gene contains a PDZ domain but no actin-binding domain. Furusawa *et al.*, *Biochem. Biophys. Res. Comm.* 270: 67-75 (2000). PDZ domains are implicated in modular protein-protein interactions, also binding to specific C-terminal sequences of membrane proteins. Gee *et al.*, *Biochem.* 39: 14638-46 (2000).

[0025] MysPDZ appears to localize in the cytoskeleton. Furusawa *et al.* show significant similarity between MysPDZ and a genomic clone from human chromosome 17; a human orthologue has been identified which contains an ATP/GTP-binding site. Nagase *et al.*, *DNA Res.* 3: 321-329 (1996). Another PDZ-domain containing protein in humans has been shown to associate with actin filaments, possibly through binding to alpha-actinin. Bauer *et al.*, *Blood* 96: 4236-45 (2000).

[0026] Given the importance of myosins in normal cellular processes and in disease, there is a need in the art for access to as yet undiscovered myosin and myosin-like genes.

SUMMARY OF THE INVENTION

[0027] These and other needs in the art are satisfied by the present invention, which provides compounds, compositions and methods based upon the identification and cloning of a novel human myosin-like protein, hGDM1P-1, which is expressed at high levels in human heart and skeletal muscle.

[0028] In a first aspect, the invention provides an isolated nucleic acid molecule selected from the group consisting of: (a) an isolated nucleic acid molecule that encodes the amino acid sequence of human genome-derived myosin-like protein, SEQ ID NO:2; (b) an isolated nucleic acid molecule that encodes a fragment of at least 10 amino acids of SEQ ID NO:2; (c) an isolated nucleic acid molecule that encodes an amino acid sequence having at least about 65% amino acid sequence identity to SEQ ID NO:2; (d) an isolated nucleic acid which hybridizes under high stringency conditions to a nucleic acid molecule comprising SEQ ID NO:1; and (e) an isolated nucleic acid complementary in sequence to the nucleic acids of (a) - (d). In one embodiment, the isolated nucleic acid comprises sequence of SEQ ID NO:1 or its complement.

[0029] In various related aspects, the invention provides a vector comprising the nucleic acids of the present invention. In certain embodiments, the vector is an expression vector. The invention further provides a host cell transformed to contain the nucleic acid molecules of the present invention, and methods for producing a polypeptide, the method comprising: culturing the transformed host cell of the present invention under conditions in which the protein encoded by the transformed nucleic acid molecule is expressed, the method optionally including a later step of isolating the expressed polypeptide.

[0030] In another aspect, the invention provides an isolated polypeptide produced by the above-described method. The invention further provides an isolated polypeptide selected from the group consisting of: (a) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2; (b) an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO:2; (c) an isolated polypeptide according to (a) or (b) in which at least 95% of deviations from the sequence of (a) or (b) are conservative substitutions; and (d) an isolated polypeptide having at least 65% amino acid sequence identity to the isolated polypeptides of (a) or (b).

[0031] In a yet further aspect, the invention provides an isolated antibody or antigen-binding fragment or derivative thereof, the binding of which can be competitively inhibited by a polypeptide of the present invention.

[0032] In other aspects, the invention provides the aforesaid nucleic acid molecules attached to a support, and in a related aspect, further provides a microarray wherein at least one probe of the array is a nucleic acid according to the invention.

[0033] In an additional aspect, the invention provides a method of identifying binding partners for a polypeptide of the present invention, the method comprising: contacting the polypeptide to a potential binding partner; and determining if the potential binding partner binds to the polypeptide. In some embodiments the contacting is performed *in vivo*.

[0034] In yet another aspect, the invention provides a method of modulating the expression of a nucleic acid of the present invention, the method comprising: administering an effective amount of an agent which modulates the expression of a nucleic acid of the present invention.

[0035] In another aspect, the invention provides a method of modulating at least one activity of a polypeptide of the present invention, the method comprising: administering an effective amount of an agent which modulates at least one activity of a polypeptide of the present invention.

[0036] The invention further provides a transgenic non-human animal modified to contain a nucleic acid molecule of the present invention. In some embodiments, the transgenic non-human animal is unable to express the polypeptide of the present invention.

[0037] The invention also provides, in a further aspect, a method of diagnosing a disease caused by mutation in hGDMLP-1, comprising: detecting the mutation in a sample of nucleic acids that derives from a subject suspected to have the disease.

[0038] The invention provides a method of diagnosing or monitoring a disease caused by altered expression of hGDMLP-1, comprising: determining the level of expression of hGDMLP-1 in a sample of nucleic acids or proteins that derives from a subject suspected to have the disease, alterations from a normal level of expression providing diagnostic and/or monitoring information.

[0039] In another aspect, the invention provides pharmaceutical compositions.

[0040] In a first such aspect, the invention provides a pharmaceutical composition comprising the nucleic acid of the present invention. In a second such aspect, the invention provides a pharmaceutical composition comprising the polypeptide of the present invention. In yet another such aspect, the invention provides a pharmaceutical composition comprising the antibody or antigen-binding fragment or derivative thereof of the present invention.

[0041] The invention further provides purified agonists and antagonists of the polypeptide of the present invention, and pharmaceutical compositions thereof.

[0042] The invention provides a method for treating or preventing a disorder associated with decreased expression or activity of hGDMLP-1, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition comprising any one of the nucleic acid, polypeptide, or agonist of the present invention. Conversely, the invention provides a method for treating or preventing a disorder associated with increased expression or activity of hGDMLP-1, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition comprising the antibody or antagonist of the present invention.

[0043] The invention further comprises diagnostic compositions, the diagnostic compositions comprising any of the nucleic acid, polypeptide, and antibody of the present invention. In certain embodiments, the diagnostic composition is suitable for *in vivo* administration.

[0044] In yet another aspect, the invention provides a method for detecting a target nucleic acid in a sample, the target being a nucleic acid of the present invention, the method comprising: a) hybridizing the sample with a probe comprising at least 30 contiguous nucleotides of a sequence complementary to the target nucleic acid in the sample under hybridization conditions sufficient to permit detectable binding of the probe to the target, and b) detecting the presence or absence, and optionally the amount, of such binding.

[0045] In a further aspect, the invention provides a fusion protein, the fusion protein comprising a polypeptide of the present invention fused to a heterologous amino acid sequence.

In certain embodiments, the heterologous amino acid sequence is a detectable moiety. In a preferred embodiment, the detectable moiety is fluorescent. In other embodiments, the heterologous amino acid sequence is an Ig Fc region.

[0046] The invention provides a method of screening for agents that modulate the expression of hGDMLP-1, the method comprising: contacting a cell or tissue sample believed to express hGDMLP-1 with a chemical or biological agent, and then comparing the amount of hGDMLP-1 expression with that of a control.

BRIEF DESCRIPTION OF THE DRAWINGS

[0047] The above and other objects and advantages of the present invention will be apparent upon consideration of the following detailed description taken in conjunction with the accompanying drawings, in which like characters refer to like parts throughout, and in which:

FIG. 1 is a screen shot of a graphical display that facilitates identification of exons in genomic sequence the expression of which meets user criteria, particularly identifying heart-specific expression of one exon of the gene of the present invention;

FIGS. 2A - 2L collectively show the sequence of the cDNA encoding the human genome-derived myosin-like protein (hGDMLP-1) of the present invention, with its predicted amino acid sequence;

FIGS. 3A - 3C collectively show an alignment of the hGDMLP-1 coding sequence of the present invention with the two closest known orthologues;

FIG. 4 shows an alignment of the amino acid sequence of hGDMLP-1 of the present invention and human myocilin;

FIG. 5 schematizes the genomic organization of the hGDMLP-1 gene, showing the relative location and size of the 44 exons of the protein and their relationship to the genomic clones (BACs) and cDNA open reading frame; and

FIG. 6 is a northern blot probed with a labeled fragment of hGDMLP-1, according to the present invention, showing muscle- and heart-specific expression of an 8 kb transcript corresponding to hGDMLP-1.

DETAILED DESCRIPTION OF THE INVENTION

[0048] Before the proteins, nucleotide sequences, and methods of the present invention are described, it is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these can be varied. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

[0049] As used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

[0050] Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Definitions

[0051] As used herein, the term "microarray" and equivalent phrase "nucleic acid microarray" refer to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed.

[0052] As so defined, the term "microarray" and phrase "nucleic acid microarray" include all the devices so called in Schena (ed.), *DNA Microarrays: A Practical Approach* (Practical Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); and Schena (ed.), *Microarray Biochip: Tools and Technology*, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of which are incorporated herein by reference in their entireties.

[0053] As so defined, the term "microarray" and phrase "nucleic acid microarray" also include substrate-bound collections of plural nucleic acids in which the nucleic acids are distributably disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000), the disclosure of which is incorporated herein by reference in its entirety; in such case, the term "microarray" and phrase "nucleic acid microarray" refer to the plurality of beads in aggregate.

[0054] As used herein with respect to a nucleic acid microarray, the term "probe" refers to the nucleic acid that is, or is intended to be, bound to the substrate. As used herein with respect to solution phase hybridization, the term "probe" refers to the nucleic acid of known sequence that is, or is intended to be, detectably labeled. In either such context, the term "target" refers to nucleic acid intended to be bound to probe by Watson-Crick complementarity. As used herein, the expression "probe comprising SEQ ID NO", and variants thereof, intends a nucleic acid probe, at least a portion of which probe has either (i) the sequence directly as given in the referenced SEQ ID NO, or (ii) a sequence complementary to the sequence as given in the referenced SEQ ID NO, the choice as between sequence directly as given and complement thereof dictated by the requirement that the probe be complementary to the desired target.

[0055] As used herein, the phrase "expression of a probe" and its linguistic variants means that the probe hybridizes detectably at high stringency to nucleic acids that derive from mRNA.

[0056] As used herein, the term "exon" refers to a nucleic acid sequence bioinformatically predicted to encode a portion of a natural protein.

[0057] As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of an exon that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF is wholly contained within its respective exon and has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

[0058] As used herein, the phrase "alternative splicing" and its linguistic equivalents includes all types of RNA processing that lead to expression of plural protein isoforms from a single gene; accordingly, the phrase "splice variant (s)" and its linguistic equivalents embraces mRNAs transcribed from a given gene that, however processed, collectively encode plural protein isoforms. For example, and by way of illustration only, splice variants can include exon insertions, exon extensions, exon truncations, exon deletions, alternatives in the 5' untranslated region ("5' UT") and alternatives in the 3' untranslated region ("3' UT"). Such 3' alternatives include, for example, differences in the site of RNA transcript cleavage and site of poly(A) addition. See, e.g., Gautheret *et al.*, *Genome Res.* 8:524-530 (1998).

[0059] As used herein, "orthologues" are separate occurrences of the same gene in multiple species. The separate occurrences have similar, albeit nonidentical, amino acid sequences, the degree of sequence similarity depending, in part, upon the evolutionary distance of the species from a common ancestor having the same gene.

[0060] As used herein, the term "paralogues" indicates separate occurrences of a gene in one species. The separate occurrences have similar, albeit nonidentical, amino acid sequences, the degree of sequence similarity depending, in part, upon the evolutionary distance from the gene duplication event giving rise to the separate occurrences.

[0061] As used herein, the term "homologues" is generic to "orthologues" and "paralogues".

[0062] Unless otherwise clear from the context, "hGDMLP-1", as used herein, refers both to the amino acid sequences of substantially purified hGDMLP-1 obtained from human beings, and additionally to orthologues thereof from any species, particularly mammalian species, including bovine, ovine, porcine, murine, rat, and equine species, from any source whether natural, synthetic, semi- synthetic, or recombinant.

[0063] The term "agonist", as used herein, refers to a molecule which, when bound to hGDMLP-1, increases or prolongs the duration of the effect of hGDMLP-1, or stimulates the function of hGDMLP-1. Agonists can include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of hGDMLP-1.

[0064] An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding hGDMLP-1. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence. "Altered" nucleic acid sequences encoding hGDMLP-1, as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent hGDMLP-1. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding hGDMLP-1, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding hGDMLP-1. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent hGDMLP-1. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of hGDMLP-1 is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

[0065] "Amino acid sequence", as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules. Fragments of hGDMLP-1 are preferably about 10 to about 15 amino acids in length and retain the biological activity or the immunological activity of hGDMLP-1. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

[0066] "Amplification", as used herein refers to the production of additional copies of a nucleic acid sequence and can be carried out techniques well known in the art, including polymerase chain reaction (PCR) (Dieffenbach, C.W. and G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Laboratory Press, Plainview, N.Y.) and rolling circle amplification (U.S. Patent Nos. 5,854,033 and 5,714,320 and international patent publications WO 97/19193 and WO 00/15779).

[0067] The term "antagonist", as used herein, refers to a molecule which, when bound to hGDMLP-1, decreases the amount or the duration of the effect of the biological or immunological activity of hGDMLP-1, or reduces the function of hGDMLP-1. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies or any other molecules which decrease the effect of hGDMLP-1.

[0068] As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fab and F(ab')₂, and derivatives thereof, such as scFV, which are capable of binding the epitopic determinant. Antibodies that bind hGDMLP-1 polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from

the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

[0069] The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

[0070] The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

[0071] The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic hGDMLP-1, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

[0072] The terms "complementary" or "complementarity", as used herein, refer to the natural duplex binding of polynucleotides under permissive salt and temperature conditions by Watson-Crick base-pairing. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

[0073] A "composition comprising a given polynucleotide sequence", as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding hGDMLP-1 or fragments thereof may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

[0074] "Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended (e.g. by RACE or anchored PCR), or has been assembled from the overlapping sequences of more than one clone using a computer program for fragment assembly. Sequences can have been extended and assembled to produce the consensus sequence.

[0075] The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:1 by northern analysis is indicative of the presence of mRNA encoding hGDMLP-1 in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

[0076] A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

[0077] The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to hGDMLP-1 or the encoded hGDMLP-1. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

[0078] The term "homology", as used herein, refers to a degree of sequence similarity.

[0079] Human artificial chromosomes (HACs) are linear microchromosomes which may contain DNA sequences of 10K to 10M basepairs in size and contain all of the elements required for stable mitotic chromosome segregation and maintenance (Harrington, J. J. et al. (1997) Nat Genet. 15:345-355).

[0080] The term "humanized antibody", as used herein, refers to nonhuman antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

[0081] The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

[0082] The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid se-

quences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

[0083] An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

[0084] The term "modulate", as used herein, refers to a change in the activity of hGDMLP-1. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of hGDMLP-1.

[0085] "Nucleic acid sequence", as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" of hGDMLP-1 are typically at least 30 nucleotides in length, at least 100 nucleotides or at least 1000 nucleotides in length, and at least 10,000 nucleotides in length.

[0086] The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or a hybridization assay, or a microarray. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers", "primers", "oligomers", and "probes", as commonly defined in the art.

[0087] "Peptide nucleic acid", PNA as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their life span in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P. E. et al. (1993) Anticancer Drug Des. 8:53-63).

[0088] The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:2" encompasses the full-length hGDMLP-1 and fragments thereof.

[0089] The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding hGDMLP-1, or fragments thereof, or hGDMLP-1 itself may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA (in solution or bound to a solid support, a tissue, a tissue print, and the like).

[0090] The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) of the protein recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

[0091] The terms "stringent conditions" or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

[0092] The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

[0093] A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

[0094] "Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating

plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

[0095] A "variant" of hGDMLP-1, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. Other variants may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software.

hGDMLP-1

[0096] The invention is based on the discovery of a new human myosin heavy chain-like protein, hGDMLP-1, the polynucleotides encoding hGDMLP-1, and the use of these compositions for the diagnosis, prevention, or treatment of disorders.

[0097] In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2, as shown in FIG. 2.

[0098] The invention also encompasses hGDMLP-1 variants. A preferred hGDMLP-1 variant is one having at least 65%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or even 99%, amino acid sequence identity to the hGDMLP-1 amino acid sequence, retaining at least one biological, immunological, or other functional characteristic or activity of hGDMLP-1. A preferred hGDMLP-1 variant is one having at least 95% amino acid sequence identity to SEQ ID NO:2.

[0099] The invention also encompasses polynucleotides which encode hGDMLP-1. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of hGDMLP-1 can be used to produce recombinant molecules which express hGDMLP-1. In a particular embodiment, the invention encompasses the polynucleotide comprising the nucleic acid sequence of SEQ ID NO:1 as shown in FIG. 2.

[0100] It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding hGDMLP-1, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring hGDMLP-1, and all such variations are to be considered as being specifically disclosed.

[0101] Although nucleotide sequences which encode hGDMLP-1 and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring hGDMLP-1 under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding hGDMLP-1 or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding hGDMLP-1 and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

[0102] The invention also encompasses production of DNA sequences, or fragments thereof, which encode hGDMLP-1 and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding hGDMLP-1 or any fragment thereof.

[0103] Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:1, under various conditions of stringency as taught in Wahl, G. M. and S. L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511).

[0104] Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (U.S. Biochemical Corp, Cleveland, Ohio), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, Ill.), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, Md.). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, Nev.), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, Mass.) and the MegaBACE™ DNA Sequencers (Molecular Dynamics™).

[0105] The nucleic acid sequences encoding hGDMLP-1 may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific

primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

[0106] Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186). The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°C-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

[0107] Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

[0108] Another method which may be used to retrieve unknown sequences is that of Parker, J. D. et al. (1991; *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto, Calif.) to walk genomic DNA. This process avoids the need to screen libraries and is useful in finding intron-exon junctions.

[0109] When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

[0110] Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. GENOTYPER and SEQUENCE NAVIGATOR, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

[0111] In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode hGDMLP-1 may be used in recombinant DNA molecules to direct expression of hGDMLP-1, fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express hGDMLP-1.

[0112] As will be understood by those of skill in the art, it may be advantageous to produce hGDMLP-1-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

[0113] The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter hGDMLP-1 encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

[0114] In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding hGDMLP-1 may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of hGDMLP-1 activity, it may be useful to encode a chimeric hGDMLP-1 protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the hGDMLP-1 encoding sequence and the heterologous protein sequence, so that hGDMLP-1 may be cleaved and purified away from the heterologous moiety.

[0115] In another embodiment, sequences encoding hGDMLP-1 may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of hGDMLP-1, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and auto-

mated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

[0116] The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of hGDMLP-1, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

[0117] In order to express a biologically active hGDMLP-1, the nucleotide sequences encoding hGDMLP-1 or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

[0118] Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding hGDMLP-1 and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.

[0119] A variety of expression vector/host systems may be utilized to contain and express sequences encoding hGDMLP-1. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

[0120] The "control elements" or "regulatory sequences" are those non-translated regions of the vector, e.g. enhancers, promoters, 5' and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla, Calif.) or PSPORTI plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding hGDMLP-1, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

[0121] In bacterial systems, a number of expression vectors may be selected depending upon the use intended for hGDMLP-1. For example, when large quantities of hGDMLP-1 are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding hGDMLP-1 may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

[0122] In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

[0123] In cases where plant expression vectors are used, the expression of sequences encoding hGDMLP-1 may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6: 307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N. Y.; pp. 191-196).

[0124] An insect system may also be used to express hGDMLP-1. For example, in one such system, Autographa

californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding hGDMLP-1 may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of hGDMLP-1 will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which hGDMLP-1 may be expressed (Engelhard, E. K. et al. (1994) *Proc. Nat. Acad. Sci.* 91:3224-3227).

[0125] In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding hGDMLP-1 may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing hGDMLP-1 in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

[0126] Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6M to 10M basepairs are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

[0127] Specific initiation signals may also be used to achieve more efficient translation of sequences encoding hGDMLP-1. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding hGDMLP-1, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125- 162).

[0128] In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

[0129] For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express hGDMLP-1 may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be expanded using tissue culture techniques appropriate to the cell type.

[0130] Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) *Cell* 22:817-23) genes which can be employed in tk⁻ or apt⁻ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047- 51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β -glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

[0131] Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding hGDMLP-1 is inserted within a marker gene sequence, transformed cells containing sequences encoding hGDMLP-1 can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding hGDMLP-1 under the control of a single promoter. Expression of the marker gene in response to induction or selection

usually indicates expression of the tandem gene as well.

[0132] Alternatively, host cells which contain the nucleic acid sequence encoding hGDMLP-1 and express hGDMLP-1 may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include

[0133] The presence of polynucleotide sequences encoding hGDMLP-1 can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding hGDMLP-1. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding hGDMLP-1 to detect transformants containing DNA or RNA encoding hGDMLP-1.

[0134] A variety of protocols for detecting and measuring the expression of hGDMLP-1, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on hGDMLP-1 is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

[0135] A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding hGDMLP-1 include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding hGDMLP-1, or any fragments thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Amersham Pharmacia Biotech, Uppsala, Sweden; Promega, Madison, Wisconsin; and U.S. Biochemical Corp., Cleveland, Ohio). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

[0136] Host cells transformed with nucleotide sequences encoding hGDMLP-1 may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode hGDMLP-1 may be designed to contain signal sequences which direct secretion of hGDMLP-1 through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding hGDMLP-1 to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor Xa or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and hGDMLP-1 may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing hGDMLP-1 and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying hGDMLP-1 from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453).

[0137] In addition to recombinant production, fragments of hGDMLP-1 may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of hGDMLP-1 may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Therapeutics

[0138] In one embodiment, hGDMLP-1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a heart or skeletal muscle disorder.

[0139] In another embodiment, a vector capable of expressing hGDMLP-1, or a fragment or a derivative thereof, may also be administered to a subject to treat or prevent a heart or skeletal muscle disorder.

[0140] In another embodiment, an agonist which modulates the activity of hGDMLP-1 may also be administered to a subject to treat or prevent a heart or skeletal muscle disorder.

[0141] In another embodiment, an antagonist of hGDMLP-1 may be administered to a subject to treat or prevent a heart or skeletal muscle disorder.

[0142] In another embodiment, an antibody which specifically binds hGDMLP-1 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express hGDMLP-1.

[0143] In another embodiment, a vector expressing the complement of the polynucleotide encoding hGDMLP-1 may be administered to a subject to treat or prevent a heart or skeletal muscle disorder.

[0144] In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

[0145] An antagonist of hGDMLP-1 may be produced using methods which are generally known in the art. In particular, purified hGDMLP-1 may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind hGDMLP-1.

[0146] Antibodies to hGDMLP-1 may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

[0147] For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with hGDMLP-1 or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette- Guerin) and *Corynebacterium parvum* are especially preferable.

[0148] It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to hGDMLP-1 have an amino acid sequence consisting of at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of hGDMLP-1 amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

[0149] Monoclonal antibodies to hGDMLP-1 may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV- hybridoma technique (Kohler, G. et al. (1975) *Nature* 256: 495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R. J. et al. (1983) *Proc. Natl. Acad. Sci.* 80: 2026-2030; Cole, S. P. et al. (1984) *Mol. Cell Biol.* 62:109-120).

[0150] In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M. S. et al. (1984) *Nature* 312: 604-608; Takeda, S. et al. (1985) *Nature* 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce hGDMLP-1-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:11120-3).

[0151] Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833- 3837; Winter, G. et al. (1991) *Nature* 349:293-299).

[0152] Antibody fragments which contain specific binding sites for hGDMLP-1 may also be generated. For example, such fragments include, but are not limited to, the F(ab')₂ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) *Science* 254:1275-1281).

[0153] Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between hGDMLP-1 and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering hGDMLP-1 epitopes is preferred, but a competitive binding assay

may also be employed (Maddox, *supra*).

[0154] In another embodiment of the invention, the polynucleotides encoding hGDMLP-1, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding hGDMLP-1 may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding hGDMLP-1. Thus, complementary molecules or fragments may be used to modulate hGDMLP-1 activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding hGDMLP-1.

[0155] Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding hGDMLP-1. These techniques are described both in Sambrook et al. (*supra*) and in Ausubel et al. (*supra*).

[0156] Genes encoding hGDMLP-1 can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes hGDMLP-1. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

[0157] As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions of the gene encoding hGDMLP-1 (signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In: Huber, B. E. and B. I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, N.Y.). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

[0158] Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding hGDMLP-1.

[0159] Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

[0160] Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding hGDMLP-1. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

[0161] RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

[0162] Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino polymers (Goldman, C. K. et al. (1997) *Nature Biotechnology* 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

[0163] Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

[0164] An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of hGDMLP-1, antibodies to hGDMLP-1, mimetics, agonists, antagonists, or inhibitors of hGDMLP-1. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

[0165] The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

[0166] In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

[0167] Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

[0168] Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

[0169] Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

[0170] Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

[0171] Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

[0172] For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

[0173] The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

[0174] The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

[0175] After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of hGDMLP-1, such labeling would include amount, frequency, and method of administration.

[0176] Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingre-

dients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

[0177] For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

[0178] A therapeutically effective dose refers to that amount of active ingredient, for example hGDMLP-1 or fragments thereof, antibodies of hGDMLP-1, agonists, antagonists or inhibitors of hGDMLP-1, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

[0179] The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

[0180] Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Diagnostics

[0181] In another embodiment, antibodies which specifically bind hGDMLP-1 may be used for the diagnosis of conditions or diseases characterized by expression of hGDMLP-1, or in assays to monitor patients being treated with hGDMLP-1, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for hGDMLP-1 include methods which utilize the antibody and a label to detect hGDMLP-1 in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

[0182] A variety of protocols including ELISA, RIA, and FACS for measuring hGDMLP-1 are known in the art and provide a basis for diagnosing altered or abnormal levels of hGDMLP-1 expression. Normal or standard values for hGDMLP-1 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to hGDMLP-1 under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of hGDMLP-1 expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

[0183] In another embodiment of the invention, the polynucleotides encoding hGDMLP-1 may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of hGDMLP-1 may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of hGDMLP-1, and to monitor regulation of hGDMLP-1 levels during therapeutic intervention.

[0184] In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding hGDMLP-1 or closely related molecules, may be used to identify nucleic acid sequences which encode hGDMLP-1. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding hGDMLP-1, alleles, or related sequences.

[0185] Probes may also be used for the detection of related sequences, and should preferably contain at least 50%

of the nucleotides from any of the hGDMLP-1 encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:1, or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring hGDMLP-1 (SEQ ID NO:3).

[0186] Means for producing specific hybridization probes for DNAs encoding hGDMLP-1 include the cloning of nucleic acid sequences encoding hGDMLP-1 or hGDMLP-1 derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ³²P or ³⁵S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

[0187] Polynucleotide sequences encoding hGDMLP-1 may be used for the diagnosis of a disorder associated with expression of hGDMLP-1, in particular heart and skeletal muscle disorders.

[0188] The polynucleotide sequences encoding hGDMLP-1 may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays or microarrays utilizing fluids or tissues from patient biopsies to detect altered hGDMLP-1 expression. Such qualitative or quantitative methods are well known in the art.

[0189] The nucleotide sequences encoding hGDMLP-1 may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding hGDMLP-1 in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

[0190] In order to provide a basis for the diagnosis of disease associated with expression of hGDMLP-1, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes hGDMLP-1, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

[0191] Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

[0192] With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

[0193] Additional diagnostic uses for oligonucleotides designed from the sequences encoding hGDMLP-1 may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation and another with antisense, employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

[0194] Methods which may also be used to quantitate the expression of hGDMLP-1 include radiolabeling or biotinylation nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229:236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

[0195] In further embodiments, an oligonucleotide derived from any of the polynucleotide sequences described herein may be used as a target in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information will be useful in determining gene function, understanding the genetic basis of disease, diagnosing disease, and in developing and monitoring the activity of therapeutic agents (Heller, R. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-55).

[0196] In one embodiment, the microarray is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference.

[0197] The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray, it may be preferable to use oligonucleotides which are only 7-10 nucleotides in length. The microarray may contain oligonucleotides which cover the known 5' or 3' sequence, sequential oligonucleotides which cover the full length sequence, or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell type, developmental or disease state.

[0198] In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5', or more preferably, at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray. The "pairs" will be identical, except for one nucleotide which preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

[0199] In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

[0200] In order to conduct sample analysis using a microarray, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray so that the probe sequences hybridize to complementary oligonucleotides of the microarray. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, mutations, variants, or polymorphisms among samples.

[0201] In another embodiment of the invention, the nucleic acid sequences which encode hGDMPL-1 may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) Blood Rev. 7:127-134, and Trask, B. J. (1991) Trends Genet. 7:149-154.

[0202] Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at Online Mendelian Inheritance in Man (OMIM). Correlation between the location of the gene encoding hGDMPL-1 on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

[0203] In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number

or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) *Nature* 336: 577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

[0204] In another embodiment of the invention, hGDMLP-1, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between hGDMLP-1 and the agent being tested, may be measured.

[0205] Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to hGDMLP-1 large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with hGDMLP-1, or fragments thereof, and washed. Bound hGDMLP-1 is then detected by methods well known in the art. Purified hGDMLP-1 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

[0206] In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding hGDMLP-1 specifically compete with a test compound for binding hGDMLP-1. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with hGDMLP-1.

[0207] In additional embodiments, the nucleotide sequences which encode hGDMLP-1 may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

[0208] The following examples are provided by way of illustration and not by way of limitation.

EXAMPLE 1

Identification and Characterization of a cDNA Encoding a Novel Human Genome-Derived Myosin-Like Gene

[0209] In order to identify novel human genes that predispose or contribute to diseases caused by perturbations of the cellular contractile apparatus, we undertook to mine human genomic DNA for exons particularly expressed in human cardiac and/or skeletal muscle, looking in particular for those that are part of genes that have proven refractory to prior identification efforts based upon cloning and/or sequencing of expressed transcripts.

[0210] Briefly, bioinformatic algorithms were applied to human genomic sequence data to identify putative exons. Each of the predicted exons was amplified from genomic DNA, typically centering the putative coding sequence within a larger amplicon that included flanking noncoding sequence. These genome-derived single exon probes were arrayed on a solid support and expression of the bioinformatically predicted exons assessed through a series of simultaneous two-color hybridizations to the genome-derived single exon microarrays. The approach and procedures are further described in detail in Penn *et al.*, "Mining the Human Genome using Microarrays of Open Reading Frames," *Nature Genetics* 26:315-318 (2000); commonly owned and copending U.S. patent application no. _____ ("Methods and Apparatus for Predicting, Confirming, and Displaying Functional information Derived from Genomic Sequence," attorney docket no. MDhMORF-1, filed January 29, 2001); and commonly owned and copending U.S. provisional patent application nos. 60/207,456, filed May 26, 2000 ("Human Genome-derived Single Exon Nucleic Acid Probes Useful for Gene Expression Analysis by Microarray"), the disclosures of which are incorporated herein by reference in their entirety.

[0211] Using a graphical display particularly designed to facilitate computerized query of the resulting exon-specific expression data, as further described in commonly owned and copending U.S. patent application no. _____ ("Methods and Apparatus for Predicting, Confirming, and Displaying Functional information Derived from Genomic Sequence," attorney docket no. MDhMORF-1, filed January 29, 2001), an exon was identified that is expressed at high levels in human heart but in none of nine other tested tissues or cell types.

[0212] FIG. 1 is a screen shot of the graphical display in which the heart-specific exon is, for purpose of exemplification, centered in the view pane. A black line, demarcated at regular intervals in megabases ("Mb"), represents a continuous stretch of human chromosome 22 ("genome sequence line"). Below the line, in a horizontal field labeled "amp", are blue rectangles that identify regions of chromosome 22 that were, in one or another experiment, amplified from genomic DNA. The blue rectangles are aligned with the genome sequence line to show the starting and ending

nucleotides of the amplicons.

[0213] Moving downward in the view pane, horizontal fields are presented that show expression measured for various of the amplicons in human heart ("HEA"), brain ("BRA"), adult liver ("ADU"), HeLa cells ("HEL"), lung ("LUN"), fetal liver ("FET"), HBL100 cells ("HBL"), bone marrow ("BON"), BT4 cells ("BT4"), and placenta ("PLA"). As further described in commonly owned and copending U.S. patent application no. _____ ("Methods and Apparatus for Predicting, Confirming, and Displaying Functional information Derived from Genomic Sequence," attorney docket no. MD-hMORF-1, filed January 29, 2001), the expression relative to control (here a pool of message from 10 tissues) is shown in shades of red and green, with green indicating expression greater than control.

[0214] As can be seen in FIG. 1, the exon centered at about 11.1052 Mb is expressed at levels above control solely (among the tested tissues and cell types) in human heart, meeting our inclusion criteria for further study.

[0215] Standard RACE (rapid amplification of cDNA ends) and RT-PCR (reverse transcription polymerase chain reaction) techniques were used to obtain overlapping cDNA clones that collectively span 8,166 nucleotides and appear to contain the entire coding region of the gene. RT-PCR and RACE techniques are described, *inter alia*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998 (ISBN: 1881299147); Schaefer, *Anal. Biochem.* 20:227(2):255-73 (1995); Ausubel *et al.* (eds.), Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 4th edition (April 1999), John Wiley & Sons (ISBN: 047132938X) and Sambrook *et al.* (eds.), Molecular Cloning: A Laboratory Manual (3rd ed.), Cold Spring Harbor Laboratory Press (2000) (ISBN: 0879695773), the disclosures of which are incorporated herein by reference in their entirety. The cDNA was sequenced on both strands using a MegaBace™ sequencer (Molecular Dynamics, Inc., Sunnyvale, CA, USA).

[0216] FIGS. 2A - 2L collectively present the sequence of the cDNA [SEQ ID NO:1] and its translation [SEQ ID NO:2]. For reasons set forth below, we have denominated the novel gene hGDMLP-1.

[0217] The hGDMLP-1 cDNA spans 8166 nucleotides and contains an open reading frame from nucleotide 170 through and including nt 7876, predicting a protein of 2568 amino acids (285.3 kD absent post-translational modification). The clone appears full length, with the reading frame opening with a methionine and terminating with a stop codon before a 3' poly-A tail.

[0218] Structurally, the N-terminal 600 amino acids, as well as the C-terminal 1100 amino acids (1400-2568) are quite hydrophilic, with fully one third of the amino acids of the first 480 residues strongly acidic or basic.

[0219] Motif searches using Pfam (<http://pfam.wustl.edu/>), SMART (<http://smart.embl-heidelberg.de/>), and PROSITE pattern and profile databases (<http://www.expasy.ch/prosite/>), identified several known domains. A myosin head (also denominated myosin motor, large ATPase) domain is encoded by amino acid residues 566 - 1335; a myosin tail domain is located at residues 1422 - 1919. Given the presence of a myosin head and a myosin tail, we have denominated the gene "human genome-derived myosin-like protein-1" ("hGDMLP-1").

[0220] In addition to myosin head and myosin tail domains, SMART and Pfam both additionally identified an IQ domain at residues 1336-1358. The IQ domain is a short calmodulin-binding motif containing conserved Ile and Gin residues, suggesting interaction of hGDMLP-1 with calmodulin, which is known to serve as the light chain for myosins I and V.

[0221] The sequence of the hGDMLP-1 cDNA was used a BLAST query into the GenBank nr and dbEst databases. The nr database includes all non-redundant GenBank coding sequence translations, sequences derived from the 3-dimensional structures in the Brookhaven Protein Data Bank (PDB), sequences from SwissProt, sequences from the protein information resource (PIR), and sequences from protein research foundation (PRF). The dbEst (database of expressed sequence tags) includes ESTs, short, single pass read cDNA (mRNA) sequences, and cDNA sequences from differential display experiments and RACE experiments.

[0222] Although a number of ESTs showed strong sequence similarity, including identity, with short regions of the hGDMLP-1 coding sequence, hGDMLP-1 has not previously been identified as a full length clone. Two full length genes prior-accessioned into GenBank do, however, appear to be closely related.

[0223] The first is a murine myosin gene containing a PDZ domain (mysPDZ) (GenBank accession no. BAA93660), further described by Furusawa *et al.*, "Isolation of a novel PDZ-containing myosin from hematopoietic supportive bone marrow stromal cell lines," *Biochem. Biophys. Res. Commun.* 270(1):67-75 (2000), the disclosure of which is incorporated herein by reference in its entirety. In the region of greatest similarity — residues 477 - 2270 of hGDMLP-1 — hGDMLP-1 and BAA93660 show 40% identity and 60% similarity at the amino acid level, with similarity scored according to the NCBI BLAST standard BLOSUM matrix (Henikoff *et al.*, *Proc. Natl. Acad. Sci. USA* 89:10915-10919 (1992); Henikoff *et al.*, *Proteins* 17:49-61 (1993)). This region includes both the hGDMLP myosin head domain and myosin tail domain.

[0224] hGDMLP-1 is also highly similar across the same region (residues 477 - 2270 of hGDMLP-1) to the human orthologue of the murine PDZ-containing myosin in mouse (accession no. BAA13206), showing 40% identity and 61% similarity to hGDMLP-1 at the amino acid level.

[0225] FIGS. 3A - 3C collectively show a multiple sequence alignment of hGDMLP-1 (amino acid residues 505-2249)

with BAA93660 (murine PDZ-containing myosin-like gene, residues 341-2035) and BAA13206 (human myosin heavy chain-like gene containing P-loop, residues 1-1581). Gaps are shown within a sequence as one or more dashes (-). In the consensus line underlying the three aligned sequences, "*" indicates residues that are identical or conserved in all sequences in the alignment, ":" indicates conserved substitutions, and "." = indicates semi-conserved substitutions.

[0226] Further directed comparison of hGDMLP-1 with 10 other proteins previously identified to be myosin or myosin-like confirms that hGDMLP-1 is most similar in sequence to mysPDZ. Significant similarity was also detected to human (but not murine) myocilin (myoC). FIG. 4 shows pairwise alignment of hGDMLP-1 to human myocilin (myoC). As shown, in the region of maximal similarity in amino acid sequence, 28% of residues are identical; 54% of residues are similar, with similarity scored according to the BLOSUM 62 matrix.

[0227] Both mysPDZ and myoC are cytoskeleton-associated proteins. MyoC is associated with glaucoma, has a secreted form, and is found in association with cilia in photoreceptor cells. Based upon these sequence similarities, hGDMLP-1 is a myosin-like protein that is cytoskeleton-associated.

[0228] Separately querying GenBank with sequence drawn from the hydrophilic N-terminus and C-terminus of hGDMLP-1, the N-terminal region (112 - 515) is seen to have some sequence similarity to collagen alpha and other hydrophilic proteins (such as cylicin and neurofilament triplet H protein), whereas the C-terminal end of the protein is unique.

EXAMPLE 2

Identification and Characterization of a Genomic Region Encoding hGDMLP-1

[0229] BLAST query of genomic sequence identified four BACs that constitute the minimum set of clones encompassing the cDNA sequence (FIG. 5).

[0230] Based upon the known origin of the four BACs (AL022329, AL080245, Z98949 and AL079300) that encompass hGDMLP-1, the gene can be mapped to human chromosome 22q11.2, about 1 Mb distal to the 22q11 low copy repeat region.

[0231] Comparison of the cDNA and genomic sequences identified 44 exons, occupying a 288 kb region on chromosome 22q11.2. Exon organization is listed in Table 1.

Table 1 hGDMLP-1 Exon Structure			
Exon no.	cDNA range	genomic range	BAC accession
1	1-60	192976-193035	AL022329.9
2	61-208	211726-211873	
3	209-367	213973-214131	
4	367-1681	218857-220170	
5	1682-1748	220852-220918	
6	1749-1861	206-318	AL080245.14
7	1862-2038	1668-1844	
8	2039-2237	6917-7115	
9	2238-2380	9390-9532	
10	2381-2481	11068-11168	
11	2482-2545	14762-14825	

Table 1
hGDMLP-1 Exon Structure

Exon no.	cDNA range	genomic range	BAC accession
12	2546-2690	27287-27431	
13	2691-2864	52839-53012	
14	2865-2955	55743-55833	
15	2956-3148	58110-58302	
16	3149-3229	62251-62331	
17	3230-3380	560-710	Z98949.1
18	3381-3540	9002-9161	
19	3541-3724	11367-11550	
20	3725-3947	12697-12919	
21	3948-4057	16737-16846	
22	4058-4123	33587-33652	
23	4124-4252	39553-39681	
24	4253-4396	41456-41599	
25	4397-4486	43226-43315	
26	4487-4606	56023-56142	
27	4607-4715	56571-56679	
28	4716-4840	60423-60547	
29	4841-4995	63574-63728	
30	4996-5119	67880-68003	
31	5120-5320	68898-69098	
32	5321-5431	73589-73699	
33	5431-5536	76213-76317	
34	5537-5689	86524-86676	
35	5690-5803	111403-111516	
36	5804-5920	112978-113094	
37	5921-6004	115633-115716	
38	6005-6142	117552-117689	
39	6143-6328	120445-120630	
40	6329-6459	157629-157759	
41	6460-6504	168531-168575	
42	6505-6642	169984-170121	

Table 1
hGDMLP-1 Exon Structure

Exon no.	cDNA range	genomic range	BAC accession
43	6643-7888	18298-19543	AL079300.11
44	7889-8090	22296-22497	

EXAMPLE 3

Expression and Transcription Control of hGDMLP-1

[0232] Exons 1, 3, 4, 6, 15, 43 and 44 of hGDMLP-1 were separately amplified and the genome-derived single-exon probes arrayed for expression analysis essentially as described in Penn *et al.*, "Mining the Human Genome using Microarrays of Open Reading Frames," *Nature Genetics* 26:315-318 (2000) and commonly owned and copending U. S. patent application no. _____ ("Methods and Apparatus for Predicting, Confirming, and Displaying Functional information Derived from Genomic Sequence," attorney docket no. MDhMORF-1, filed January 29, 2001), the disclosures of which are incorporated herein by reference in their entireties.

[0233] Table 2 reports the results of hybridization experiments. "Amplicon" is a unique designation. "Signal" indicates normalized signal intensity. "Ratio" is the ratio of expression relative to control. "ND" indicates "not determined".

Table 2. Expression of hGDMPLP-1 by microarray analysis.

	Amp_2534, exon 1	Amp_2541, exon 3	Amp_2540, exon 3	Amp_38, exon 4	Amp_10400, exon 4	Amp_119					
	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL
ADRENAL	0.71	-	0.94	-	0.7	1.15	4.12	-	ND	ND	ND
ADULT LIVER	0.52	1.13	0.46	-	0.59	-	6.54	-1.28	1.56	-	2.41
BONE MARROW	0.71	-	0.77	-	0.95	-	5.18	-1.29	1.77	-1.15	2.35
BRAIN	0.52	-1.06	1.28	-	0.83	-	5.62	-1.58	1.8	-1.15	1.75
BT474	ND	ND	ND	ND	ND	ND	ND	ND	1.57	-	1.51
FETAL LIVER	0.49	-	0.47	-	0.72	1.06	4.48	-1.3	1.49	-	2.47
HBL100	ND	ND	ND	ND	ND	ND	ND	ND	6.44	-	1.71
HEART	1.86	3.33	2.01	-	2.52	2.77	6.23	-1.05	2.9	1.36	3.16
HELA	0.71	-	0.72	-1.03	0.82	1.03	6.08	-1.11	1.91	1.11	2.07
KIDNEY	0.61	-	0.66	1	0.81	-	6.71	-1.09	ND	ND	ND
LUNG	0.66	-1.21	0.73	-	0.83	-	7.15	-1.02	0.99	-1.38	1.95
PLACENTA	0.46	-	0.55	2.59	0.57	-	4.14	-	2.09	1.02	2.2
PROSTATE	0.26	-	0.58	-	0.58	1.15	5.13	-1.16	ND	ND	ND
SKELETAL MUSCLE	9.61	-	3.38	5.14	4.04	-	8.12	1.04	ND	ND	ND

Table 2. Expressio

6, exon 4 Amp_9715, exon 6 Amp_10252, exon 15 Amp_40, exon 43 Amp_2920, exon 44

	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL	RATIO	SIGNAL
ADRENAL	ND	ND	ND	ND	ND	ND	0.9	1.26	-	1.26	-	-
ADULT LIVER	-	2.88	-	1.28	-	0.82	-	1.3	-	1.3	-	-1.34
BONE MARROW	1.08	2.99	-1.22	1.68	-1.1	1.51	-	1.31	-	1.31	-	-1.14
BRAIN	-1.23	3.92	1.09	1.86	-1.08	0.99	1.28	1.01	-	1.01	-	-1.3
BT474	-1.63	3.35	-	2.11	-	ND	ND	ND	ND	ND	ND	ND
FETAL LIVER	-	3.14	-1.2	1.31	-1.27	0.72	1.07	1.2	-	1.2	-	-1.5
HBL100	-1.61	2.25	-1.43	1.7	-1.08	ND	ND	ND	ND	ND	ND	ND
HEART	-1.02	3.05	-1.13	1.9	1.26	5.21	-	1.97	-	1.97	-	-
HELA	-1.01	3.05	-1.17	1.36	-	0.9	-	1.13	-	1.13	-	1
KIDNEY	ND	ND	ND	ND	-	0.99	-	1.08	-	1.08	-	-1.01
LUNG	-1.05	2.59	-1.15	1.03	-1.33	0.84	-	0.81	-	0.81	-	-
PLACENTA	-1.12	3.19	-1.05	1.37	-1.23	0.71	-	1.14	-	1.14	-	-1.36
PROSTATE	ND	ND	ND	ND	ND	0.73	-	2.02	-	2.02	-	-
SKELETAL MUSCLE	ND	ND	ND	ND	ND	15.04	-	3.6	-	3.6	-	-

[0234] As can be seen in table 2, skeletal muscle and heart are the primary locations of hGDMLP-1 expression.

[0235] Heart and skeletal muscle-specific expression was further confirmed by northern blot analysis using amplicon 40 (exon 43) as probe. As shown in FIG. 6, of the 12 tissues assayed — blood, leukocytes, lung, placenta, small intestine, liver, kidney, spleen, thymus, colon, skeletal muscle, heart and brain — only skeletal muscle and heart demonstrate expression, with both having an 8 kb transcript.

[0236] Possession of the genomic sequence permitted search for promoter and other control sequences for the

hGDMPLP-1 gene.

[0237] A putative transcriptional control region, inclusive of promoter and downstream elements, was defined as 2 kb around the transcription start site, itself defined as the first nucleotide of the cDNA clone. The region, drawn from sequence of BAC AL022329, has the following sequence, where nucleotide number 1001 is the transcription start site:

```

5      tgatTTTTTTTTTTTTTTTTTgagatggagtcctctctgtcaccaggtgga
10     gtgcagtggcgcgatctcggttactgcaagctccgcctgtagctgggactac
      aggcgcccaccaccacgcccggctaattttgtttttgtatTTTtagtagagac
      ggggtttcactgtgttagccaggatggtctcaatctcctgacctcgtgatatg
15     cccacctcggttcccaaagtgccgggattacagacgtgagccaccgtgccag
      gcccaaagcactTTTTTatagcccctagtttgctgttggtctgtttcctcta
      cggcccgcacagcctgggtttgaattctggctctccatggccttgggcaagtt
20     ggttaactTTTTTgtgtctcaatttccccacttgtaaagtggggctggggaaaa
      accctacctcatagggaaacattaagtgagttcatgaaagccctcggtgtgtgg
      catgcgataaccactacctaagcataaggatatgtttacatgcactggttggtt
25     accacgacgcaccagcattcattcaagttatTTTcttccagagtctgcctgga
      atgtagaaccggccagctctcctcctcaaacagttcctgcaacctctggttat
      ctgtgtgggcagaggcgggagtggtgggtgtgtggccctgccctgggacagggg
30     tgacgagagagcatgtgcttggtgtccagagctgtccccttccctgctcagggg
      ccaccaaatagtccatgggtctctgctgcctgggtcaacggccacccgggctg
      ccgtctcctggccagacctggagctcacaagctaattTTTtagaaatgactggtc
35     gactcagctgctgctatcacattggcccagaaggagggtggggggaggacggg

```

tcggggaggggcagtaaggggaagggatgtgtgtcggtgtgtgtggagagGGT
 5 GCTCCTTCGCTCCTGCTTCCCCGGCTCGGTGGGGTCCCTTCGCGGCcaccg
 ggagacgtcttagacagtcgtggttcctgtgatctgtgtcctctttgcagaag
 gtactattgggctgggttttagacgggcaccttgcgagcggatcggcgagttt
 10 taatthttccggaatgcgctgctgggggctgcggtggcaggcctccggtctttt
 ggtctatthtttaggcctttctggtgttgagagcttccgaggcctgattcctttg
 cctgttgcccgctgctgctcactttgaaaacaagagtgtttcggtggaaggatt
 15 tgagaacaaaatgctggggcaaatggacattcgtagaatcactctcctcagg
 cccagctgttcgctaataatthtttaaaaatgtgtttagaatcatgagttaat
 ctccactgtgaatctttgcccgcagacgtcagggtttatatthtttctcaact
 accagtaatthttctgcttcactggtgttcagttatggtcggggaagagccaggc
 20 tccccctttcgtgttaacaacgctgttgagtcgaaatgttaatthttaatgttc
 thttthtaactgcaaattaagcatgtatgagtgttgggggaagggcaatggaa
 gttcatggctgaaggthttcaatthcatgccaccttccttgagctctgtctgag
 25 ggaagggtgtcttatctggcaccgccgatagcaggcatttctctgaatgaagaa
 gattcacagggtcccagtggtcccatagtggggcagaagggggctatgctttga
 catggagctctctcccaggagaggagaagacctcaccagctacaggcctctgg
 30 gctgctgccttcctgcagctthttccatggaggaggccaggagtggggaggccc
 ggaatggggagcccagtcggctcccagtcggctccaagctgtccagctctggcc
 tcaaccgcgaagtgtggggatctgggcttctatcctgtagggcctggaatacg
 35 acgaggctgttactthtcactthtggtggtgtccaaaaaga [SEQ ID NO:3]

40 **[0238]** Using PROSCAN,
 (<http://bimas.dcrn.nih.gov/molbio/proscan/>), no significant promoter was identified in the putative promoter region. How-
 ever, transcription factor binding sites were identified using a web site at <http://motif.genome.ad.jp/>, including a SRY
 (sex-determining region Y gene product) binding sites (134..140 bp, with numbering according to SEQ ID NO:3), and
 45 a couple of MZF1 (myeloid zinc finger 1, a negative regulator of CD34 and c-myb, and responsive to retinoic acid)
 binding sites (1843..1850 bp and 403-395 bp).

EXAMPLE 3

hGDMLP-1 Disease Associations

50 **[0239]** Diseases that map to the hGDMLP-1 chromosomal region are shown in Table 3.

Table 3
Co-mapping Diseases

OMIM No.	name	map location
115470	cat eye syndrome	22q11
217095	conotruncal cardiac anomalies	22q11
192430	DiGeorge syndrome (2); Velocardiofacial syndrome	22q11
604364	Epilepsy, partial, with variable foci	22q11-q12
600850	Schizophrenia susceptibility locus, chromosome 22-related	22q11-q13
155100	May-Hegglin anomaly	22q11.2
153640	Fechtner syndrome	22q11.2
605249	Sebastian syndrome	22q11.2
603622	Deafness, autosomal dominant 17	22q11.2

Table 3
Co-mapping Diseases

OMIM No.	name	map location
145410	Opitz G syndrome, type II	22q11.2
601547	Cataract, cerulean, type 2	22q11.2-q12.2

[0240] A number of these diseases have physiologic characteristics consistent with alteration of a myosin-like gene, especially a myosin-like gene particularly expressed in human heart.

[0241] Thus, the iris alterations that give cat eye syndrome (CES) its name are frequently associated with heart malformations. A small supernumerary chromosome (smaller than chromosome 21) is present, frequently has 2 centromeres, is bisatellited, and represents an inv dup(22)(q11). No gene candidates have been identified as yet.

[0242] Individuals with 22q11 deletion syndrome (del 22q11) have a range of findings, including congenital heart disease (74% of patients), particularly conotruncal malformations (tetralogy of Fallot, interrupted aortic arch, and truncus arteriosus); palatal abnormalities (69%), particularly velopharyngeal incompetence (VPI), submucosal cleft palate, and cleft palate; characteristic facial features (present in the majority of individuals); and learning difficulties (70 - 90%).

[0243] DiGeorge syndrome (DGS), most cases of which result from a deletion of chromosome 22q11.2, is characterized by hypocalcemia arising from parathyroid hypoplasia, thymic hypoplasia, and outflow tract defects of the heart. This syndrome has previously been considered to be a contiguous gene deletion, but one or a few genes in the identified locus may play a major role in the pathogenesis of the syndrome's characteristic phenotypic features. Several genes are lost in the characteristic deletion, but definitive identification of the causative genes is lacking.

[0244] Several disorders have been tentatively linked to another myosin-like gene that maps to 22q11, MYH9, although some of these associations are inferential and may be due instead to mutations in hGDMLP-1.

[0245] For example, the autosomal dominant giant-platelet disorders May-Hegglin anomaly, Fechtner syndrome, and Sebastian syndrome share the triad of thrombocytopenia, large platelets, and characteristic leukocyte inclusions. All three disorders may be allelic since May-Hegglin anomaly and Fechtner syndrome map to an overlapping region of 480 kb on chromosome 22. MYH9 was recently identified as a possible candidate gene, since MYH9 is expressed in platelets and is upregulated during granulocyte differentiation, but association of the disease with MYH9 is inferential. Kelley *et al.*, *Nature Genetics* 26:106-108 (2000).

[0246] Other disorders that map to 22q11 may result from defects of hGDMLP-1, although primary disease manifestations are not in cardiac or skeletal muscle.

[0247] For example, Xiong *et al.*, *Am. J. Hum. Genet.* 65:1698-1710 (1999), mapped the disease locus for a familial partial epilepsy syndrome with variable foci to an interval on 22q11-q12. The syndrome is characterized by mostly nocturnal seizures arising from frontal, temporal, and occasionally occipital epileptic foci. The syndrome is inherited as an autosomal dominant trait with incomplete penetrance.

[0248] Several studies have suggested that there may be a susceptibility locus for schizophrenia located on chromosome 22. Gill *et al.*, *Am. J. Med. Genet.* 67:40-5 (1996). Additionally, Pulver *et al.* (1994) reported a higher incidence of schizophrenia among patients with velocardiofacial syndrome, which is localized to 22q11. Candidate genes have not yet been identified.

[0249] Opitz G syndrome, Opitz *et al.*, *Birth Defects Orig. Art. Ser.* 2: 95-101 (1969), is characterized by hypertelorism or telecanthus; laryngotracheoesophageal cleft; clefts of lip, palate, and uvula; swallowing difficulty and hoarse cry; genitourinary defects, especially hypospadias in males and splayed labia majora in females; mental retardation; and congenital heart defects. Opitz Syndrome may also be characterized by additional abnormalities, including partial or complete closure of the anal opening (imperforate anus); hypoplasia or agenesis of the corpus callosum; renal abnormalities; heart defects; or mental retardation. Although Opitz G syndrome has been mapped to 22q11.2, no specific genes have been associated with the disease. See, e.g., Robin *et al.*, *Am. J. Med. Genet.* 62: 305-317 (1996).

[0250] Congenital cataracts are responsible for between 10 and 30 % of all cases of blindness in children. One of these congenital forms is cerulean cataract which has been mapped to 22q11.2-22q12.2. The beta-crystallin genes have been identified as possible candidate genes, but no definitive association has been shown.

[0251] All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. By their citation of various references in this document, applicants do not admit that any particular reference is "prior art" to their invention.

[0252] While specific examples have been provided, the above description is illustrative and not restrictive. Any one or more of the features of the previously described embodiments can be combined in any manner with one or more features of any other embodiments in the present invention. Furthermore, many variations of the invention will become apparent to those skilled in the art upon review of the specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

EP 1 158 049 A1

<110> AEOMICA INC.

<120> MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

5 <130> PB 0105-EP

<140> EP 01112637.2-2105

<141> 2001-05-24

10 <150> US 60/207,456

<151> 2000-05-26

<150> US 60/236,359

<151> 2000-09-27

15 <150> PCT/US01/00666

<151> 2001-01-30

<150> PCT/US01/00667

<151> 2001-01-30

20 <150> PCT/US01/00664

<151> 2001-01-30

<150> PCT/US01/00669

<151> 2001-01-30

25 <150> PCT/US01/00665

<151> 2001-01-30

<150> PCT/US01/00668

<151> 2001-01-30

30 <150> PCT/US01/00663

<151> 2001-01-30

<150> PCT/US01/00662

<151> 2001-01-30

35 <150> PCT/US01/00661

<151> 2001-01-30

<150> PCT/US01/00670

<151> 2001-01-30

<150> US 60/234,687

<151> 2000-09-21

45 <150> US 60/266,860

<151> 2001-02-05

<160> 3

<170> Aeomica Sequence Listing Engine

50 <210> 1

<211> 8117

<212> DNA

<213> Homo sapiens

55 <220>

<221> CDS
 <222> 170-7876

```

5      <400> 1
      caccgcgggga gacgtcttag acagtcgtgg ttctctgtgat ctgtgtcctc tttgcagaag 60
      ttctgtgtcc atctcatgtg ctgctgtgtg ctgtaaagcc tcattccgtg ctgtctggca 120
      ggaagctcca tctcatctca tcatctcacg gccctggcac tgccctcagca tggccatctc 180
      atcacgcctc gccctgtggg agcagaagat tcgggaagag gacaagagcc ctccaccatc 240
      ctgcgccctt cctcttttct ctgtcatccc aggggggcttc attagcaaac tgggtccgggg 300
10     gactgaaaaa gagggcaagg aagcgagaca gaggaagcag ttagctgtcg cctctccaga 360
      acgagagatc ccagaaatct ccatacagcca acccaacagc aagtcacagca gtggcaccag 420
      atctggaagc cagcagatct ctcaggacga ccagtcgaag tctcctggga gctcagacat 480
      tctgggcaag gagagcgagg ggtcccgag ccccgacct gagcagatga caagcatcaa 540
      tgggtgagaag gcccgaggag tgggctccag tgcgacacca accaaaaaga ctgtcccctt 600
15     caagaggggc gtgaggaggg gtgatgtgtt gttgatgggt gccaaactgg acccgactc 660
      agccaagcca gagaagactc atccccatga cgcgccctt tgcaagacct ctccccccgc 720
      cacagatact ggaaaggaaa agaaagggga gacctctagg actccttgtg gctcccaggc 780
      cagcaccgag atcttggccc cgaaagctga gaagaccgg actgggggtc ttggggaccc 840
      aggccaaaga actgtggcac tgaaaaaagg cgaggagggt caaagcatag tggggaaggg 900
      gcttgggacc cccaagacca cagagctgaa agaggctgag cccagggca aagacaggca 960
20     ggggaccagg ccccaagccc aagggcccg cgaggggggt cgaccaggga aagcagagaa 1020
      ggaggagca gagccacaa acacggtgga aaaggggaat gtctctaagg acgtaggag 1080
      tgaagggaag cacgtaaggc ccaaatccc tgggagaaag tggggagggt tcctgggaag 1140
      aaggagtaag tgggacggtc cccagaataa gaaggacaaa gaaggggtgc tcttaagtaa 1200
      ggcagagaag acaggtgagc ctcagaccca gatggagaag acaagccaag tgcagggcga 1260
      gttgggggac gatctgagaa tgggggagaa agcagggtgag cttcggagca cgactgggaa 1320
25     ggcaggtgag tcttgggata agaaggaaaa gatggggcaa cccagggta agtcgggaa 1380
      cgcaggtgaa gctcggagtc agacagagaa gggctgtgaa gccccaaagg aggtgagcac 1440
      aatgggtggag tcgccagcag ctctgggaa gggaggctgg ccaggaagcc gtgggcaggga 1500
      agcagaggag cctgtctcaa gagcagggtga tggggctggt gccctggaga cagagctgga 1560
      aggaccagc cagcctgtct tggagaagga tgcagaaaag cctcggatac ggaaggagaa 1620
30     ccaagacggg ccagccccgc aggaggaggg caaaggaggc cagagcagag actcagacca 1680
      ggctcctgag gacagatggt atgaggcaga gaaagtctgg ctggctcaga aggatggatt 1740
      tactcttgct acggtgctaa agccagatga gggaacagca gacctgccag caggaagggt 1800
      gagactttgc attgatgtg acaaaacat cactgagtg gatgaggagc atgtccatcg 1860
      ggccaacctt cctgagctgg accagggtcg ggacctggcc tctctcatca gtgtcaacga 1920
      atccagtgtc ctgaacacgc ttctgcagcg ctacaaagct cagctgtctc acacctgcac 1980
35     agggcctgat ctgattgtcc tccagccccg ggggccctcg gtgccttctg cagggaagggt 2040
      gcccaagggc cgccgggatg gcctgcctgc ccacattggc tccatggcac agcgggcata 2100
      ctgggcgctg ctgaaccagc ggagagacca gagcattgtg gccctgggcc ggagtggcgc 2160
      tgggaagacc acctgctgtg agcaggctct ggaacacctg gtggggatgg caggcagtg 2220
      ggatggcagg attctagtg agaagatccg agccaccttc actgtcctcc gggccttcgg 2280
      ctctgtgtcc atggcccaca gccgcagtg cacccggttc tccatggtga tgtcgtgga 2340
40     cttcaacgct acaggccgca tcacagctgc tcagctccag acaatgcttt tggagaagag 2400
      ccgcgtggca cggcagcccg aaggggaaaag taacttctcg gttttctccc agatgctggc 2460
      tggattggac ttggatctca ggacggagct gaacctgcac cagatggcag atagcagctc 2520
      ctttggcatg ggcgtgtggt ccaagcctga agataaacag aaggcggcag ctgcctttgc 2580
      ccagctccag ggtgccatgg agatgctggg catctcagag agcgagcagc gggctgtttg 2640
45     gcgggtcctg gcagccatct accacctgg cgcgcgggg gcctgcaaag tgggtcggaa 2700
      gcagttcatg aggtttgagt gggcaaaacta gcagctgag gccctgggct gcgagtatga 2760
      ggagctgaac acggccacct tcaagcacca ccttcgacag atcatccagc aaatgacgtt 2820
      tgggccaagc cgatggggcc tcgaggatga ggaaaccagc tcagggtca agatgacagg 2880
      agtggactgt gtggagggga tggcctcggg cctgtaccag gaactctttg cggctgtggt 2940
      ctactcatc aacagatcct ttctctcca ccactctcc atggcctcca tcatggtggt 3000
50     ggactctcca ggcttcaga acccccgga ccagggcaag gaccggcgcc ccacctttga 3060
      ggagctgtgc cacaactacg cccatgagcg cctgcagctg ctgttctacc agcggacctt 3120
      tgtctccacg ctacagcgat atcaagagga aggtgttct gtgcagttt accctccgga 3180
      cccctccca gggaccacg tggctgttgt ggcataaaat cctctcagc aggtccgctt 3240
      accagctgga ggaggtgcc aggatgccag aggccttttc tgggtcttag atgaggaagt 3300
      ccatgtagag ggctccagtg acagtgtggt gctcgagcgt ctgtgtgcta ctttcgagaa 3360
55     gaaaggagct gggactgaag ggtcctctgc cctgcggacc tgtgagcagc ccctccagtg 3420

```


	tgagattttc	caccagttgg	gatgggaccc	tgtgcggtac	gacctcacgg	gctggctcca	3480
	cagagccaag	cccaacctct	cggccctgga	tgacccccag	gtcctgcaac	agtcaaaaag	3540
	agaggagctg	cggagtctat	tccaggcccg	ggccaagctg	cctcctgtgt	gccgggctgt	3600
5	ggcaggcctg	gagggcacct	cccagcaggc	cctgcagagg	agccgcatgg	tgaggaggac	3660
	ctttgccagc	agccttgccg	cggtgaggag	gaaaagcccc	tgctcccaga	tcaagctgca	3720
	gatggatgcg	ctgaccagca	tgatcaaaaag	gtcccggctg	cactttatcc	actgcctggt	3780
	accaaaccct	gtggtggaaa	gcaggagtgg	gcaggaatct	ccaccaccac	cgcagcctgg	3840
	tagagacaag	cctggggcag	gtggacctct	ggccctggat	atcccagcac	tgagggtcca	3900
	gcttgctggg	ttccacatcc	tggaggctct	gcgtctgcat	aggacaggct	atgctgacca	3960
10	catggggctc	actcgcttcc	gccggcaatt	ccagggtgctg	gacgctccac	tcctgaagaa	4020
	gctcatgtcg	acctccgagg	gaatagatga	aaggaaaggcc	gtggaggagc	tcctggagac	4080
	cctggatctg	gaaaagaagg	cgggtggctgt	ggggcacagc	caagtttttc	tcaaggcagg	4140
	tgtgatctcc	aggcttgaga	agcagcgaga	gaagctggta	tctcagagca	tcgttctctt	4200
	ccaggcggct	tgcaagggct	ttctgtctcg	ccaggaatcc	aagaagctga	agattcgccg	4260
	actggctgca	cagtgcattc	agaagaatgt	ggctgtgttc	ctcgcagtca	aggactggcc	4320
15	atggtggcag	ctgcttggtt	ccctccagcc	tctacttagt	gccaccattg	gaactgagca	4380
	gtcccgagcc	aaggaggagg	agcttacaac	gctaagacgg	aagctagaaa	aatcagagaa	4440
	gttgcggaat	gaactccggc	agaacacaga	tctgctagaa	agcaagattg	ctgacttgac	4500
	ctctgacctt	gccgatgagc	gcttcaaagg	tgatgtggcc	tgccagggtg	tggagagtga	4560
	gcgggacag	cggctacagg	ccttccggga	gggtccaggag	ctcaagagca	agcatgaaca	4620
20	agtccagaaa	aaactgggag	atgtgaataa	acagttggaa	gaagcccagc	agaaaattca	4680
	gttgaatgac	ttggaaaagg	atcccactgg	aggagcagac	gagtggcaga	tgcgcttcga	4740
	ctgtgctcag	atggagaacg	agttcctcag	aaagcgtctg	cagcaatgcg	aggagaggct	4800
	ggactcggag	ctgacagcca	ggaaagagct	ggagcaaaaag	cttggggagt	tgcaaagtgc	4860
	ttatgacggg	ccaagaaga	tggctcacca	actgaagagg	aagtgccacc	atcttacctg	4920
	tgaccttgag	gatacctgcg	tcctgctaga	gaaccaacaa	agtcgaaacc	atgagctgga	4980
25	gaagaagcag	aagaagtttg	acctgcagct	ggcccaggcc	ctaggtgagt	cagtgtttga	5040
	gaagggctct	cgtgagaaag	tgacccagga	gaacaccagt	gtccggtggg	agctaggcca	5100
	gcttcagcag	cagctgaagc	aaaaggagca	ggaagcctca	cagctgaagc	agcagggtga	5160
	gatgctacag	gaccataaac	gggagctgct	gggggtcacc	tctctggggg	aaaattgcgt	5220
	tgctggcttg	aaggagaggc	tctggaagtt	ggaatccagc	gcccttgagc	aacagaaaat	5280
30	ccagagccag	caggaaaaca	ccatcaagca	gctggagcag	ctccgccagc	ggtttgagct	5340
	ggagatcgag	cggatgaagc	agatgcacca	gaagaccgtg	gaggaccagg	aggaggaact	5400
	ggaggatgtc	cgtcagtoct	cccagaagcg	gcttcatcag	ctggaaatgc	agctggagca	5460
	agagtatgaa	gagaagcaga	tggctcctcca	tgagaagcaa	gatttggaag	gcttgatcgg	5520
	aacctctctg	gaccagattg	gccatcgggg	ctttgatgtg	gagaagcgac	ttcggagaga	5580
	cctcaggagg	acacatgcac	tgttgctaga	cgtgcagctc	cttctgggca	ccatggagga	5640
35	tggaagaca	tcagtcagca	aggaggagct	ggagaaagtg	cacagccagc	tggagcagag	5700
	tgaagccaag	tgtgaggagg	ccttgaagac	gcagaagggtg	ctcacagcgg	acctggagag	5760
	catgcacagc	gagctggaga	acatgacgcg	gaacaagagc	ctggtggatg	agcagctgta	5820
	caggctgcag	tttgagaagg	cggacctcct	gaagcgcatc	gatgaggacc	aggatgacct	5880
	gaatgagctg	atgcagaagc	acaaggacct	cattgctcag	tctgctgctg	acattgggca	5940
	gatccaagaa	ctgcagctgc	agctggagga	agccaagaag	gagaagcaca	agctacaaga	6000
40	acaactgcag	gtggctcaga	tgcgcatcga	gtacctggaa	cagtccaccg	tggatcgagc	6060
	catcgtcagc	aggcaggagg	cggctcatctg	tgacctagag	aacaagacag	agttccagaa	6120
	ggtgcagatt	aagagatttg	aggtcctggt	gatccggctt	cgggacagcc	tgatcaagat	6180
	gggggaggag	ctttcacagg	cggccacctc	caggtcccag	cagcggggaga	gcagccagta	6240
	ctaccagcgg	cgcttggaa	agctgaaggc	cgacatggaa	gagctgggtg	agcgggaggc	6300
45	agagggcagc	cggcggtgca	tggagctgga	gaagtacgtg	gaggaacttg	cagcagtgag	6360
	gcaaaccctc	cagacagacc	tggagacatc	cattccggcg	attgccgacc	tgagggtgct	6420
	cttggaaagaa	gtggcatcca	gtgacagtga	tactgagagt	gtccagacgg	cagtggattg	6480
	tggcagcagc	ggccgaaaag	agatggataa	cgtctccatc	ctcagctccc	agccagaggg	6540
	cagcctgcag	tcctgggtga	gctgtactct	gtccctggcc	acagatacta	tgaggactcc	6600
	ttctcgacag	tcagccacca	gcagccgcat	cctcagcccc	aggataaacg	aagaggctgg	6660
50	ggacactgag	aggacccagt	cggcattggc	actgagcaga	gcccggtcca	ccaatgtcca	6720
	cagcaagacc	tcaggagaca	agcctgtttc	tccccacttt	gtccgcgggc	aaaagtactg	6780
	tcattttggg	gacggcgagg	tgcttgccgt	ccagagaaaag	tccacagaga	gattagaacc	6840
	tgtctcctct	cccctggctt	ctcggagtac	aaatacatcc	ccgctgtoga	gggaaaagct	6900
	gcccagtcct	tcagcggccc	tctcgagttt	ctgggaaggg	ctccggaggga	agagagccca	6960
55	gagaggccag	gggtccacgc	tgggcctaga	ggactggccc	actctcccca	tttaccagac	7020
	gactggggcc	tcacactaa	ggagggcgag	ggctggcagt	gacgagggaa	acctctcgct	7080

EP 1 158 049 A1

```

gaggggttggg gcaaagtcac ccctggaaat cgaagggggcc gctgggtggtc tcttgagggtc 7140
caccagcctc aaatgcatct cttcagacgg tggtgggggc acaaccctac tccccgaaaa 7200
gtcgaaaacc caattcagtt cctgcgagtc cctcttagaa tccagaccga gcatggggag 7260
5 aaaactgagc tctccgacca caccagggga catgctgttg tcgcccacac tgcgtcctcg 7320
gaggcgggtg ctggagtcct ctgtggacga tgcgggctgt ccagacctg gaaaggagcc 7380
gcttgttttc cagaaccgcc agtttgccca cctgatggag gaacctctag gcagtgacct 7440
attcagctgg aaactcccaa gcctcgacta cgaacgcaag accaaagtgg acttcgatga 7500
cttctcccca gctatccgga agccccagac acctacctcc ttggctggat cagccaaagg 7560
10 tgggcaagac ggttcacagc gttcaagcat ccactttgaa acggaagagg ctaaccgttc 7620
ctttctctcg gggatcaaga ccattttgaa gaagagcccg gagcccaagg aggatcccg 7680
tcacctgtct gactcgtcct catctcccg ctccatcggt tccttcaaaa gtgctgacag 7740
catcaaaagt cgaccaggaa tcccacgact tgcgggtgac ggtggcgagc gaacgtcccc 7800
cgagcggaga gagccaggga cggggaggaa agacgacgat gttgcgagca taatgaagaa 7860
atacctccag aagtaggaac cagttcagcc tccttgaagc tgcccttgaa gacttcccga 7920
15 ctctacaata acttgagac agagagactg gccaggcctc cccggtggcc agagccagcc 7980
agcatggcca ccctcaagag gcgagatgag cccacagagg catatcctgc ggggatgctg 8040
ggctcccagt gtggttgcc tgaacaaaat aaagtgttga ctctgggca aaaaaaaaaa 8100
aaaaaaaaa aaaaaaa 8117

```

```

<210> 2
<211> 2568
<212> PRT
<213> Homo sapiens

```

```

<400> 2
25 Met Ala Ile Ser Ser Arg Leu Ala Leu Trp Glu Gln Lys Ile Arg Glu
1 5 10 15

Glu Asp Lys Ser Pro Pro Pro Ser Ser Pro Pro Pro Leu Phe Ser Val
20 25 30

30 Ile Pro Gly Gly Phe Ile Lys Gln Leu Val Arg Gly Thr Glu Lys Glu
35 40 45

Ala Lys Glu Ala Arg Gln Arg Lys Gln Leu Ala Val Ala Ser Pro Glu
50 55 60

35 Arg Glu Ile Pro Glu Ile Ser Ile Ser Gln Pro Asn Ser Lys Ser Ser
65 70 75 80

Ser Gly Thr Arg Ser Gly Ser Gln Gln Ile Ser Gln Asp Asp Gln Ser
85 90 95

40 Ser Ser Pro Gly Ser Ser Asp Ile Leu Gly Lys Glu Ser Glu Gly Ser
100 105 110

Arg Ser Pro Asp Pro Glu Gln Met Thr Ser Ile Asn Gly Glu Lys Ala
115 120 125

45 Gln Glu Leu Gly Ser Ser Ala Thr Pro Thr Lys Lys Thr Val Pro Phe
130 135 140

Lys Arg Gly Val Arg Arg Gly Asp Val Leu Leu Met Val Ala Lys Leu
145 150 155 160

50 Asp Pro Asp Ser Ala Lys Pro Glu Lys Thr His Pro His Asp Ala Pro
165 170 175

55 Pro Cys Lys Thr Ser Pro Pro Ala Thr Asp Thr Gly Lys Glu Lys Lys
180 185 190

```

EP 1 158 049 A1

	Gly	Glu	Thr	Ser	Arg	Thr	Pro	Cys	Gly	Ser	Gln	Ala	Ser	Thr	Glu	Ile	
				195				200					205				
5	Leu	Ala	Pro	Lys	Ala	Glu	Lys	Thr	Arg	Thr	Gly	Gly	Leu	Gly	Asp	Pro	
		210					215					220					
	Gly	Gln	Gly	Thr	Val	Ala	Leu	Lys	Lys	Gly	Glu	Glu	Gly	Gln	Ser	Ile	
	225					230					235					240	
10	Val	Gly	Lys	Gly	Leu	Gly	Thr	Pro	Lys	Thr	Thr	Glu	Leu	Lys	Glu	Ala	
					245					250					255		
	Glu	Pro	Gln	Gly	Lys	Asp	Arg	Gln	Gly	Thr	Arg	Pro	Gln	Ala	Gln	Gly	
				260					265					270			
15	Pro	Gly	Glu	Gly	Val	Arg	Pro	Gly	Lys	Ala	Glu	Lys	Glu	Gly	Ala	Glu	
		275						280					285				
	Pro	Thr	Asn	Thr	Val	Glu	Lys	Gly	Asn	Val	Ser	Lys	Asp	Val	Gly	Ser	
20		290					295					300					
	Glu	Gly	Lys	His	Val	Arg	Pro	Gln	Ile	Pro	Gly	Arg	Lys	Trp	Gly	Gly	
	305					310					315					320	
	Phe	Leu	Gly	Arg	Arg	Ser	Lys	Trp	Asp	Gly	Pro	Gln	Asn	Lys	Lys	Asp	
25					325					330					335		
	Lys	Glu	Gly	Val	Leu	Leu	Ser	Lys	Ala	Glu	Lys	Thr	Gly	Glu	Pro	Gln	
				340					345					350			
	Thr	Gln	Met	Glu	Lys	Thr	Ser	Gln	Val	Gln	Gly	Glu	Leu	Gly	Asp	Asp	
30			355					360					365				
	Leu	Arg	Met	Gly	Glu	Lys	Ala	Gly	Glu	Leu	Arg	Ser	Thr	Thr	Gly	Lys	
		370					375					380					
	Ala	Gly	Glu	Ser	Trp	Asp	Lys	Lys	Glu	Lys	Met	Gly	Gln	Pro	Gln	Gly	
35		385				390					395					400	
	Lys	Ser	Gly	Asn	Ala	Gly	Glu	Ala	Arg	Ser	Gln	Thr	Glu	Lys	Gly	Cys	
				405					410						415		
	Glu	Ala	Pro	Lys	Glu	Val	Ser	Thr	Met	Val	Glu	Ser	Pro	Ala	Ala	Pro	
40				420				425						430			
	Gly	Lys	Gly	Gly	Trp	Pro	Gly	Ser	Arg	Gly	Gln	Glu	Ala	Glu	Glu	Pro	
		435					440						445				
	Cys	Ser	Arg	Ala	Gly	Asp	Gly	Ala	Gly	Ala	Leu	Glu	Thr	Glu	Leu	Glu	
45		450					455					460					
	Gly	Pro	Ser	Gln	Pro	Ala	Leu	Glu	Lys	Asp	Ala	Glu	Arg	Pro	Arg	Ile	
	465					470					475					480	
50	Arg	Lys	Glu	Asn	Gln	Asp	Gly	Pro	Ala	Pro	Gln	Glu	Glu	Gly	Lys	Gly	
				485						490					495		
	Gly	Gln	Ser	Arg	Asp	Ser	Asp	Gln	Ala	Pro	Glu	Asp	Arg	Trp	Tyr	Glu	
				500					505					510			
55	Ala	Glu	Lys	Val	Trp	Leu	Ala	Gln	Lys	Asp	Gly	Phe	Thr	Leu	Ala	Thr	

EP 1 158 049 A1

	515	520	525
5	Val Leu Lys Pro Asp Glu Gly Thr Ala Asp Leu Pro Ala Gly Arg Val 530 535 540		
	Arg Leu Cys Ile Asp Ala Asp Lys Thr Ile Thr Glu Val Asp Glu Glu 545 550 555 560		
10	His Val His Arg Ala Asn Pro Pro Glu Leu Asp Gln Val Glu Asp Leu 565 570 575		
	Ala Ser Leu Ile Ser Val Asn Glu Ser Ser Val Leu Asn Thr Leu Leu 580 585 590		
15	Gln Arg Tyr Lys Ala Gln Leu Leu His Thr Cys Thr Gly Pro Asp Leu 595 600 605		
	Ile Val Leu Gln Pro Arg Gly Pro Ser Val Pro Ser Ala Gly Lys Val 610 615 620		
20	Pro Lys Gly Arg Arg Asp Gly Leu Pro Ala His Ile Gly Ser Met Ala 625 630 635 640		
	Gln Arg Ala Tyr Trp Ala Leu Leu Asn Gln Arg Arg Asp Gln Ser Ile 645 650 655		
25	Val Ala Leu Gly Arg Ser Gly Ala Gly Lys Thr Thr Cys Cys Glu Gln 660 665 670		
	Val Leu Glu His Leu Val Gly Met Ala Gly Ser Val Asp Gly Arg Val 675 680 685		
30	Ser Val Glu Lys Ile Arg Ala Thr Phe Thr Val Leu Arg Ala Phe Gly 690 695 700		
	Ser Val Ser Met Ala His Ser Arg Ser Ala Thr Arg Phe Ser Met Val 705 710 715 720		
35	Met Ser Leu Asp Phe Asn Ala Thr Gly Arg Ile Thr Ala Ala Gln Leu 725 730 735		
	Gln Thr Met Leu Leu Glu Lys Ser Arg Val Ala Arg Gln Pro Glu Gly 740 745 750		
40	Glu Ser Asn Phe Leu Val Phe Ser Gln Met Leu Ala Gly Leu Asp Leu 755 760 765		
	Asp Leu Arg Thr Glu Leu Asn Leu His Gln Met Ala Asp Ser Ser Ser 770 775 780		
45	Phe Gly Met Gly Val Trp Ser Lys Pro Glu Asp Lys Gln Lys Ala Ala 785 790 795 800		
50	Ala Ala Phe Ala Gln Leu Gln Gly Ala Met Glu Met Leu Gly Ile Ser 805 810 815		
	Glu Ser Glu Gln Arg Ala Val Trp Arg Val Leu Ala Ala Ile Tyr His 820 825 830		
55	Leu Gly Ala Ala Gly Ala Cys Lys Val Gly Arg Lys Gln Phe Met Arg 835 840 845		

EP 1 158 049 A1

	Phe	Glu	Trp	Ala	Asn	Tyr	Ala	Ala	Glu	Ala	Leu	Gly	Cys	Glu	Tyr	Glu	
	850						855					860					
5	Glu	Leu	Asn	Thr	Ala	Thr	Phe	Lys	His	His	Leu	Arg	Gln	Ile	Ile	Gln	
	865					870					875					880	
	Gln	Met	Thr	Phe	Gly	Pro	Ser	Arg	Trp	Gly	Leu	Glu	Asp	Glu	Glu	Thr	
10					885					890					895		
	Ser	Ser	Gly	Leu	Lys	Met	Thr	Gly	Val	Asp	Cys	Val	Glu	Gly	Met	Ala	
				900					905					910			
	Ser	Gly	Leu	Tyr	Gln	Glu	Leu	Phe	Ala	Ala	Val	Val	Ser	Leu	Ile	Asn	
15			915					920					925				
	Arg	Ser	Phe	Ser	Ser	His	His	Leu	Ser	Met	Ala	Ser	Ile	Met	Val	Val	
	930					935						940					
	Asp	Ser	Pro	Gly	Phe	Gln	Asn	Pro	Arg	His	Gln	Gly	Lys	Asp	Arg	Ala	
20	945				950						955					960	
	Ala	Thr	Phe	Glu	Glu	Leu	Cys	His	Asn	Tyr	Ala	His	Glu	Arg	Leu	Gln	
				965					970						975		
	Leu	Leu	Phe	Tyr	Gln	Arg	Thr	Phe	Val	Ser	Thr	Leu	Gln	Arg	Tyr	Gln	
25				980					985					990			
	Glu	Glu	Gly	Val	Pro	Val	Gln	Phe	Asp	Leu	Pro	Asp	Pro	Ser	Pro	Gly	
			995					1000					1005				
	Thr	Thr	Val	Ala	Val	Val	Asp	Gln	Asn	Pro	Ser	Gln	Gln	Val	Arg	Leu	
30			1010				1015					1020					
	Pro	Ala	Gly	Gly	Gly	Ala	Gln	Asp	Ala	Arg	Gly	Leu	Phe	Trp	Val	Leu	
	1025					1030					1035					1040	
	Asp	Glu	Glu	Val	His	Val	Glu	Gly	Ser	Ser	Asp	Ser	Val	Val	Leu	Glu	
35				1045					1050						1055		
	Arg	Leu	Cys	Ala	Thr	Phe	Glu	Lys	Lys	Gly	Ala	Gly	Thr	Glu	Gly	Ser	
				1060					1065					1070			
40	Ser	Ala	Leu	Arg	Thr	Cys	Glu	Gln	Pro	Leu	Gln	Cys	Glu	Ile	Phe	His	
			1075				1080						1085				
	Gln	Leu	Gly	Trp	Asp	Pro	Val	Arg	Tyr	Asp	Leu	Thr	Gly	Trp	Leu	His	
	1090					1095						1100					
45	Arg	Ala	Lys	Pro	Asn	Leu	Ser	Ala	Leu	Asp	Ala	Pro	Gln	Val	Leu	Gln	
	1105				1110						1115					1120	
	Gln	Ser	Lys	Arg	Glu	Glu	Leu	Arg	Ser	Leu	Phe	Gln	Ala	Arg	Ala	Lys	
50				1125						1130					1135		
	Leu	Pro	Pro	Val	Cys	Arg	Ala	Val	Ala	Gly	Leu	Glu	Gly	Thr	Ser	Gln	
				1140					1145					1150			
	Gln	Ala	Leu	Gln	Arg	Ser	Arg	Met	Val	Arg	Arg	Thr	Phe	Ala	Ser	Ser	
55			1155				1160						1165				

EP 1 158 049 A1

	Leu	Ala	Ala	Val	Arg	Arg	Lys	Ala	Pro	Cys	Ser	Gln	Ile	Lys	Leu	Gln	
	1170						1175					1180					
5	Met	Asp	Ala	Leu	Thr	Ser	Met	Ile	Lys	Arg	Ser	Arg	Leu	His	Phe	Ile	
	1185						1190				1195					1200	
	His	Cys	Leu	Val	Pro	Asn	Pro	Val	Val	Glu	Ser	Arg	Ser	Gly	Gln	Glu	
					1205					1210					1215		
10	Ser	Pro	Pro	Pro	Pro	Gln	Pro	Gly	Arg	Asp	Lys	Pro	Gly	Ala	Gly	Gly	
					1220					1225					1230		
	Pro	Leu	Ala	Leu	Asp	Ile	Pro	Ala	Leu	Arg	Val	Gln	Leu	Ala	Gly	Phe	
					1235					1240					1245		
15	His	Ile	Leu	Glu	Ala	Leu	Arg	Leu	His	Arg	Thr	Gly	Tyr	Ala	Asp	His	
	1250						1255					1260					
	Met	Gly	Leu	Thr	Arg	Phe	Arg	Arg	Gln	Phe	Gln	Val	Leu	Asp	Ala	Pro	
20	1265					1270					1275					1280	
	Leu	Leu	Lys	Lys	Leu	Met	Ser	Thr	Ser	Glu	Gly	Ile	Asp	Glu	Arg	Lys	
					1285					1290					1295		
	Ala	Val	Glu	Glu	Leu	Leu	Glu	Thr	Leu	Asp	Leu	Glu	Lys	Lys	Ala	Val	
25					1300					1305					1310		
	Ala	Val	Gly	His	Ser	Gln	Val	Phe	Leu	Lys	Ala	Gly	Val	Ile	Ser	Arg	
					1315					1320					1325		
	Leu	Glu	Lys	Gln	Arg	Glu	Lys	Leu	Val	Ser	Gln	Ser	Ile	Val	Leu	Phe	
30	1330						1335					1340					
	Gln	Ala	Ala	Cys	Lys	Gly	Phe	Leu	Ser	Arg	Gln	Glu	Phe	Lys	Lys	Leu	
	1345					1350					1355					1360	
35	Lys	Ile	Arg	Arg	Leu	Ala	Ala	Gln	Cys	Ile	Gln	Lys	Asn	Val	Ala	Val	
					1365					1370					1375		
	Phe	Leu	Ala	Val	Lys	Asp	Trp	Pro	Trp	Trp	Gln	Leu	Leu	Gly	Ser	Leu	
					1380					1385					1390		
40	Gln	Pro	Leu	Leu	Ser	Ala	Thr	Ile	Gly	Thr	Glu	Gln	Leu	Arg	Ala	Lys	
					1395					1400					1405		
	Glu	Glu	Glu	Leu	Thr	Thr	Leu	Arg	Arg	Lys	Leu	Glu	Lys	Ser	Glu	Lys	
					1410					1415					1420		
45	Leu	Arg	Asn	Glu	Leu	Arg	Gln	Asn	Thr	Asp	Leu	Leu	Glu	Ser	Lys	Ile	
	1425					1430					1435					1440	
	Ala	Asp	Leu	Thr	Ser	Asp	Leu	Ala	Asp	Glu	Arg	Phe	Lys	Gly	Asp	Val	
					1445					1450					1455		
50	Ala	Cys	Gln	Val	Leu	Glu	Ser	Glu	Arg	Ala	Glu	Arg	Leu	Gln	Ala	Phe	
					1460					1465					1470		
	Arg	Glu	Val	Gln	Glu	Leu	Lys	Ser	Lys	His	Glu	Gln	Val	Gln	Lys	Lys	
55					1475					1480					1485		
	Leu	Gly	Asp	Val	Asn	Lys	Gln	Leu	Glu	Glu	Ala	Gln	Gln	Lys	Ile	Gln	

EP 1 158 049 A1

	1490	1495	1500
5	Leu Asn Asp Leu Glu Arg Asp Pro Thr Gly Gly Ala Asp Glu Trp Gln 1505 1510 1515 1520		
	Met Arg Phe Asp Cys Ala Gln Met Glu Asn Glu Phe Leu Arg Lys Arg 1525 1530 1535		
10	Leu Gln Gln Cys Glu Glu Arg Leu Asp Ser Glu Leu Thr Ala Arg Lys 1540 1545 1550		
	Glu Leu Glu Gln Lys Leu Gly Glu Leu Gln Ser Ala Tyr Asp Gly Ala 1555 1560 1565		
15	Lys Lys Met Ala His Gln Leu Lys Arg Lys Cys His His Leu Thr Cys 1570 1575 1580		
	Asp Leu Glu Asp Thr Cys Val Leu Leu Glu Asn Gln Gln Ser Arg Asn 1585 1590 1595 1600		
20	His Glu Leu Glu Lys Lys Gln Lys Lys Phe Asp Leu Gln Leu Ala Gln 1605 1610 1615		
	Ala Leu Gly Glu Ser Val Phe Glu Lys Gly Leu Arg Glu Lys Val Thr 1620 1625 1630		
25	Gln Glu Asn Thr Ser Val Arg Trp Glu Leu Gly Gln Leu Gln Gln Gln 1635 1640 1645		
	Leu Lys Gln Lys Glu Gln Glu Ala Ser Gln Leu Lys Gln Gln Val Glu 1650 1655 1660		
30	Met Leu Gln Asp His Lys Arg Glu Leu Leu Gly Ser Pro Ser Leu Gly 1665 1670 1675 1680		
	Glu Asn Cys Val Ala Gly Leu Lys Glu Arg Leu Trp Lys Leu Glu Ser 1685 1690 1695		
35	Ser Ala Leu Glu Gln Gln Lys Ile Gln Ser Gln Gln Glu Asn Thr Ile 1700 1705 1710		
	Lys Gln Leu Glu Gln Leu Arg Gln Arg Phe Glu Leu Glu Ile Glu Arg 1715 1720 1725		
40	Met Lys Gln Met His Gln Lys Asp Arg Glu Asp Gln Glu Glu Glu Leu 1730 1735 1740		
	Glu Asp Val Arg Gln Ser Cys Gln Lys Arg Leu His Gln Leu Glu Met 1745 1750 1755 1760		
45	Gln Leu Glu Gln Glu Tyr Glu Glu Lys Gln Met Val Leu His Glu Lys 1765 1770 1775		
	Gln Asp Leu Glu Gly Leu Ile Gly Thr Leu Cys Asp Gln Ile Gly His 1780 1785 1790		
50	Arg Asp Phe Asp Val Glu Lys Arg Leu Arg Arg Asp Leu Arg Arg Thr 1795 1800 1805		
55	His Ala Leu Leu Ser Asp Val Gln Leu Leu Leu Gly Thr Met Glu Asp 1810 1815 1820		

EP 1 158 049 A1

	Gly Lys Thr Ser Val Ser Lys Glu Glu Leu Glu Lys Val His Ser Gln	
	1825	1830 1835 1840
5	Leu Glu Gln Ser Glu Ala Lys Cys Glu Glu Ala Leu Lys Thr Gln Lys	
		1845 1850 1855
	Val Leu Thr Ala Asp Leu Glu Ser Met His Ser Glu Leu Glu Asn Met	
		1860 1865 1870
10	Thr Arg Asn Lys Ser Leu Val Asp Glu Gln Leu Tyr Arg Leu Gln Phe	
		1875 1880 1885
	Glu Lys Ala Asp Leu Leu Lys Arg Ile Asp Glu Asp Gln Asp Asp Leu	
		1890 1895 1900
15	Asn Glu Leu Met Gln Lys His Lys Asp Leu Ile Ala Gln Ser Ala Ala	
		1905 1910 1915 1920
	Asp Ile Gly Gln Ile Gln Glu Leu Gln Leu Gln Leu Glu Glu Ala Lys	
		1925 1930 1935
20	Lys Glu Lys His Lys Leu Gln Glu Gln Leu Gln Val Ala Gln Met Arg	
		1940 1945 1950
	Ile Glu Tyr Leu Glu Gln Ser Thr Val Asp Arg Ala Ile Val Ser Arg	
		1955 1960 1965
25	Gln Glu Ala Val Ile Cys Asp Leu Glu Asn Lys Thr Glu Phe Gln Lys	
		1970 1975 1980
	Val Gln Ile Lys Arg Phe Glu Val Leu Val Ile Arg Leu Arg Asp Ser	
		1985 1990 1995 2000
	Leu Ile Lys Met Gly Glu Glu Leu Ser Gln Ala Ala Thr Ser Glu Ser	
		2005 2010 2015
30	Gln Gln Arg Glu Ser Ser Gln Tyr Tyr Gln Arg Arg Leu Glu Glu Leu	
		2020 2025 2030
	Lys Ala Asp Met Glu Glu Leu Val Gln Arg Glu Ala Glu Ala Ser Arg	
		2035 2040 2045
35	Arg Cys Met Glu Leu Glu Lys Tyr Val Glu Glu Leu Ala Ala Val Arg	
		2050 2055 2060
	Gln Thr Leu Gln Thr Asp Leu Glu Thr Ser Ile Arg Arg Ile Ala Asp	
		2065 2070 2075 2080
40	Leu Gln Ala Ala Leu Glu Glu Val Ala Ser Ser Asp Ser Asp Thr Glu	
		2085 2090 2095
	Ser Val Gln Thr Ala Val Asp Cys Gly Ser Ser Gly Arg Lys Glu Met	
		2100 2105 2110
45	Asp Asn Val Ser Ile Leu Ser Ser Gln Pro Glu Gly Ser Leu Gln Ser	
		2115 2120 2125
50	Trp Leu Ser Cys Thr Leu Ser Leu Ala Thr Asp Thr Met Arg Thr Pro	
		2130 2135 2140
55		

EP 1 158 049 A1

	Ser Arg Gln Ser Ala Thr Ser Ser Arg Ile Leu Ser Pro Arg Ile Asn	
	2145	2150 2155 2160
5	Glu Glu Ala Gly Asp Thr Glu Arg Thr Gln Ser Ala Leu Ala Leu Ser	
		2165 2170 2175
	Arg Ala Arg Ser Thr Asn Val His Ser Lys Thr Ser Gly Asp Lys Pro	
		2180 2185 2190
10	Val Ser Pro His Phe Val Arg Arg Gln Lys Tyr Cys His Phe Gly Asp	
		2195 2200 2205
	Gly Glu Val Leu Ala Val Gln Arg Lys Ser Thr Glu Arg Leu Glu Pro	
		2210 2215 2220
15	Ala Ser Ser Pro Leu Ala Ser Arg Ser Thr Asn Thr Ser Pro Leu Ser	
		2225 2230 2235 2240
	Arg Glu Lys Leu Pro Ser Pro Ser Ala Ala Leu Ser Glu Phe Val Glu	
		2245 2250 2255
20	Gly Leu Arg Arg Lys Arg Ala Gln Arg Gly Gln Gly Ser Thr Leu Gly	
		2260 2265 2270
	Leu Glu Asp Trp Pro Thr Leu Pro Ile Tyr Gln Thr Thr Gly Ala Ser	
		2275 2280 2285
25	Thr Leu Arg Arg Gly Arg Ala Gly Ser Asp Glu Gly Asn Leu Ser Leu	
		2290 2295 2300
	Arg Val Gly Ala Lys Ser Pro Leu Glu Ile Glu Gly Ala Ala Gly Gly	
		2305 2310 2315 2320
30	Leu Leu Arg Ser Thr Ser Leu Lys Cys Ile Ser Ser Asp Gly Val Gly	
		2325 2330 2335
	Gly Thr Thr Leu Leu Pro Glu Lys Ser Lys Thr Gln Phe Ser Ser Cys	
		2340 2345 2350
35	Glu Ser Leu Leu Glu Ser Arg Pro Ser Met Gly Arg Lys Leu Ser Ser	
		2355 2360 2365
	Pro Thr Thr Pro Arg Asp Met Leu Leu Ser Pro Thr Leu Arg Pro Arg	
		2370 2375 2380
40	Arg Arg Cys Leu Glu Ser Ser Val Asp Asp Ala Gly Cys Pro Asp Leu	
		2385 2390 2395 2400
	Gly Lys Glu Pro Leu Val Phe Gln Asn Arg Gln Phe Ala His Leu Met	
		2405 2410 2415
45	Glu Glu Pro Leu Gly Ser Asp Pro Phe Ser Trp Lys Leu Pro Ser Leu	
		2420 2425 2430
50	Asp Tyr Glu Arg Lys Thr Lys Val Asp Phe Asp Asp Phe Leu Pro Ala	
		2435 2440 2445
	Ile Arg Lys Pro Gln Thr Pro Thr Ser Leu Ala Gly Ser Ala Lys Gly	
		2450 2455 2460
55	Gly Gln Asp Gly Ser Gln Arg Ser Ser Ile His Phe Glu Thr Glu Glu	

```

2465                2470                2475                2480
5  Ala Asn Arg Ser Phe Leu Ser Gly Ile Lys Thr Ile Leu Lys Lys Ser
    2485                2490                2495

    Pro Glu Pro Lys Glu Asp Pro Ala His Leu Ser Asp Ser Ser Ser Ser
        2500                2505                2510

10 Ser Gly Ser Ile Val Ser Phe Lys Ser Ala Asp Ser Ile Lys Ser Arg
    2515                2520                2525

    Pro Gly Ile Pro Arg Leu Ala Gly Asp Gly Gly Glu Arg Thr Ser Pro
        2530                2535                2540

15 Glu Arg Arg Glu Pro Gly Thr Gly Arg Lys Asp Asp Asp Val Ala Ser
    2545                2550                2555                2560

    Ile Met Lys Lys Tyr Leu Gln Lys
        2565

<210> 3
<211> 1000
<212> DNA
25 <213> Homo sapiens

<400> 3
tgattttttt tttttttttt gagatggagt ctctctctgt caccagggt ggagtgcagt 60
ggcgcgatct cggcttactg caagctccgc ctgtagctgg gactacaggc gccaccacc 120
acgcccggct aattttgttt ttgtattttt agtagagacg gggtttctact gtgttagcca 180
ggatggtctc aatctctctga cctcgtgata tgcccacctc ggcttcccaa agtgccggga 240
ttacagacgt gagccaccgt gccaggccca aagcactttt ttatagcccc tagtttgctg 300
ttgtgtctgt ttctctctacg gcccgcacag cctggggttg aattctggct ctccatggcc 360
ttgggcaagt tggttaactt tttgtgtctc aatttcccca cttgtaaagt ggggctgggg 420
35 aaaaacccta cctcataggg aacattaagt gagttcatga aagccctcgg tgtgtggcat 480
gcgataacca ctacctaaagc ataaggatatg ttacatgca ctggttggtt accacgacgc 540
accagcattc attcaagtta ttttcttcca gagtctgcct ggaatgtaga accggccagc 600
tctctctctc aaacagttcc tgcaacctct ggttatctgt gtgggcagag gcgggagtg 660
gggtgtgttg cctgcccctg ggacagggat gacgagagag catgtgcttg tgtccagagc 720
tgtccccctc cctgctcagg ggccacccaa tatgtccatg gtctctgctg cctgggtcaac 780
40 ggccaccccg ggctgcgctc tcttgccag acctggagct cacaagctaa ttttagaaat 840
gactggtcga ctcagctgct gctatcacat tggcccagaa ggagggtggg gggaggacgg 900
gtcggggagg ggcagtaagg ggaagggatg tgtgtcgggt tgtgtggaga ggggtgctct 960
tcgctcctgc tttcccggct cgggtggggtc ccttcgcggc 1000

```

Claims

- 50 1. An isolated nucleic acid molecule selected from the group consisting of: (a) an isolated nucleic acid molecule that encodes the amino acid sequence of SEQ ID NO:2; (b) an isolated nucleic acid molecule that encodes a fragment of at least 10 amino acids of SEQ ID NO:2; (c) an isolated nucleic acid molecule that encodes an amino acid sequence having at least about 65% amino acid sequence identity to SEQ ID NO:2; (d) an isolated nucleic acid which hybridizes under high stringency conditions to a nucleic acid molecule comprising SEQ ID NO:1; and (e) an
55 isolated nucleic acid complementary in sequence to the nucleic acids of (a) - (d).
2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule comprises sequence of SEQ ID NO:1 or its complement.

3. The isolated nucleic acid molecule of claim 2, wherein said nucleic acid molecule comprises nucleotides 170 - 7876 of SEQ ID NO:1 or the complement thereof.
- 5 4. The isolated nucleic acid molecule of any one of claims 1 - 3, wherein said nucleic acid molecule is operably linked to one or more expression control elements.
5. A replicable vector comprising an isolated nucleic acid molecule of any one of claims 1 - 4.
- 10 6. The isolated nucleic acid molecule of any of claims 1 - 3, attached to a support.
7. A host cell transformed to contain the nucleic acid molecule of any one of claims 1 - 5, or the progeny thereof.
8. A method for producing a polypeptide, the method comprising: culturing the host cell of claim 7 under conditions in which the protein encoded by said nucleic acid molecule is expressed.
- 15 9. An isolated polypeptide produced by the method of claim 8.
10. An isolated polypeptide selected from the group consisting of: (a) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2; (b) an isolated polypeptide comprising a fragment of at least 10 amino acids of SEQ ID NO:2; (c) an isolated polypeptide according to (a) or (b) in which at least 95% of deviations from the sequence of (a) or (b) are conservative substitutions; and (d) an isolated polypeptide having at least 65% amino acid sequence identity to the isolated polypeptides of (a) or (b).
- 20 11. An isolated antibody or antigen-binding fragment or derivative thereof the binding of which can be competitively inhibited by a polypeptide of any one of claims 9 or 10.
- 25 12. A method of modulating at least one activity of a polypeptide according to claim 10, the method comprising: administering an effective amount of an agent which modulates at least one activity of a polypeptide according to claim 10.
- 30 13. A method of diagnosing a disease caused by mutation in hGDMLP-1, comprising: detecting said mutation in a sample of nucleic acids that derives from a subject suspected to have said disease.
- 35 14. A method of diagnosing or monitoring a disease caused by altered expression of hGDMLP-1, comprising: determining the level of expression of hGDMLP-1 in a sample of nucleic acids or proteins that derives from a subject suspected to have said disease, alterations from a normal level of expression providing diagnostic and/or monitoring information.
- 40 15. A pharmaceutical composition comprising the nucleic acid of any one of claims 1 - 5 and a pharmaceutically acceptable excipient.
16. A pharmaceutical composition comprising the polypeptide of claim 9 or claim 10 and a pharmaceutically acceptable excipient.
- 45 17. A pharmaceutical composition comprising the antibody or antigen-binding fragment or derivative thereof of claim 11 and a pharmaceutically acceptable excipient.
18. A diagnostic composition comprising the nucleic acid of any one of claims 1 - 5, said nucleic acid being detectably labeled.
- 50 19. A diagnostic composition comprising the polypeptide of claim 9 or claim 10, said polypeptide being detectably labeled.
- 55 20. A diagnostic composition comprising the antibody or antigen-binding fragment or derivative thereof of claim 11.
21. The diagnostic composition of claim 20, wherein said antibody or antigen-binding fragment or derivative thereof is detectably labeled.

22. A microarray wherein at least one probe of said array is a nucleic acid according to any one of claims 1 - 5.

23. A method for detecting a target nucleic acid in a sample, said target being a nucleic acid of any one of claims 1 - 5, the method comprising:

- a) hybridizing the sample with a probe comprising at least 30 contiguous nucleotides of a sequence complementary to said target nucleic acid in said sample under hybridization conditions sufficient to permit detectable binding of said probe to said target, and
- b) detecting the presence or absence, and optionally the amount, of said binding.

24. A fusion protein, said fusion protein comprising a polypeptide of claim 10 fused to a heterologous amino acid sequence.

25. The fusion protein of claim 24, wherein said heterologous amino acid sequence is a detectable moiety.

26. A method of screening for agents that modulate the expression of hGDMLP-1, the method comprising:

- contacting a cell or tissue sample believed to express hGDMLP-1 with a chemical or biological agent, and then comparing the amount of hGDMLP-1 expression with that of a control.



SEQUENCE OF hGDMLP-1

cDNA: 1 - 8117 [SEQ ID NO:1]

CDS: 170 - 7876 [SEQ ID NO:2]

```

caccgcgggagacgtcttagacagtcgtggttcctgtgatctgtgtcctctt 52
tgcagaagttctgtgtccatctcatgtgctgcgtgtgtctgttaaagcctcat 104
tccgtgctgtctggcaggaagctccatctcatctcatctcacggccctg 156

      M   A   I   S   S   R   L   A   L           9
gcactgcctcagc ATG GCC ATC TCA TCA CGC CTC GCC CTG 196

W   E   Q   K   I   R   E   E   D   K   S   P   P   22
TGG GAG CAG AAG ATT CGG GAA GAG GAC AAG AGC CCT CCA 235

P   S   S   P   P   P   L   F   S   V   I   P   G   35
CCA TCC TCG CCC CCT CCT CTT TTC TCT GTC ATC CCA GGG 274

G   F   I   K   Q   L   V   R   G   T   E   K   E   48
GGC TTC ATT AAG CAA CTG GTC CGG GGG ACT GAA AAA GAG 313

A   K   E   A   R   Q   R   K   Q   L   A   V   A   61
GCC AAG GAA GCG AGA CAG AGG AAG CAG TTA GCT GTC GCC 352

S   P   E   R   E   I   P   E   I   S   I   S   Q   74
TCT CCA GAA CGA GAG ATC CCA GAA ATT TCC ATC AGC CAA 391

P   N   S   K   S   S   S   G   T   R   S   G   S   87
CCC AAC AGC AAG TCC AGC AGT GGC ACC AGA TCT GGA AGC 430

Q   Q   I   S   Q   D   D   Q   S   S   S   P   G   100
CAG CAG ATC TCT CAG GAC GAC CAG TCA AGC TCT CCT GGG 469

S   S   D   I   L   G   K   E   S   E   G   S   R   113
AGC TCA GAC ATT CTG GGC AAG GAG AGC GAG GGG TCC CGC 508

S   P   D   P   E   Q   M   T   S   I   N   G   E   126
AGC CCC GAC CCT GAG CAG ATG ACA AGC ATC AAT GGT GAG 547

K   A   Q   E   L   G   S   S   A   T   P   T   K   139
AAG GCC CAG GAG CTG GGC TCC AGT GCG ACA CCA ACC AAA 586

K   T   V   P   F   K   R   G   V   R   R   G   D   152
AAG ACT GTC CCC TTC AAG AGG GGC GTG AGG AGG GGT GAT 625

V   L   L   M   V   A   K   L   D   P   D   S   A   165
GTG TTG TTG ATG GTG GCC AAG CTG GAC CCG GAC TCA GCC 664

K   P   E   K   T   H   P   H   D   A   P   P   C   178

```

FIG. 2A

AAG	CCA	GAG	AAG	ACT	CAT	CCC	CAT	GAC	GCC	CCC	CCT	TGC	703
K	T	S	P	P	A	T	D	T	G	K	E	K	191
AAG	ACC	TCT	CCC	CCC	GCC	ACA	GAT	ACT	GGA	AAG	GAA	AAG	742
K	G	E	T	S	R	T	P	C	G	S	Q	A	204
AAA	GGG	GAG	ACC	TCT	AGG	ACT	CCT	TGT	GGC	TCC	CAG	GCC	781
S	T	E	I	L	A	P	K	A	E	K	T	R	217
AGC	ACC	GAG	ATC	TTG	GCC	CCG	AAA	GCT	GAG	AAG	ACC	CGG	820
T	G	G	L	G	D	P	G	Q	G	T	V	A	230
ACT	GGG	GGT	CTT	GGG	GAC	CCA	GGC	CAA	GGA	ACT	GTG	GCA	859
L	K	K	G	E	E	G	Q	S	I	V	G	K	243
CTG	AAA	AAA	GGC	GAG	GAG	GGT	CAA	AGC	ATA	GTG	GGG	AAG	898
G	L	G	T	P	K	T	T	E	L	K	E	A	256
GGG	CTT	GGG	ACC	CCC	AAG	ACC	ACA	GAG	CTG	AAA	GAG	GCT	937
E	P	Q	G	K	D	R	Q	G	T	R	P	Q	269
GAG	CCC	CAG	GGC	AAA	GAC	AGG	CAG	GGG	ACC	AGG	CCC	CAA	976
A	Q	G	P	G	E	G	V	R	P	G	K	A	282
GCC	CAA	GGG	CCC	GGC	GAG	GGG	GTG	CGA	CCA	GGG	AAA	GCA	1015
E	K	E	G	A	E	P	T	N	T	V	E	K	295
GAG	AAG	GAG	GGA	GCA	GAG	CCC	ACA	AAC	ACG	GTG	GAA	AAG	1054
G	N	V	S	K	D	V	G	S	E	G	K	H	308
GGG	AAT	GTC	TCT	AAG	GAC	GTA	GGG	AGT	GAA	GGG	AAG	CAC	1093
V	R	P	Q	I	P	G	R	K	W	G	G	F	321
GTA	AGG	CCC	CAA	ATC	CCT	GGG	AGA	AAG	TGG	GGA	GGT	TTC	1132
L	G	R	R	S	K	W	D	G	P	Q	N	K	334
CTG	GGA	AGA	AGG	AGT	AAG	TGG	GAC	GGT	CCC	CAG	AAT	AAG	1171
K	D	K	E	G	V	L	L	S	K	A	E	K	347
AAG	GAC	AAA	GAA	GGG	GTG	CTC	TTA	AGT	AAG	GCA	GAG	AAG	1210
T	G	E	P	Q	T	Q	M	E	K	T	S	Q	360
ACA	GGT	GAG	CCT	CAG	ACC	CAG	ATG	GAG	AAG	ACA	AGC	CAA	1249
V	Q	G	E	L	G	D	D	L	R	M	G	E	373
GTG	CAG	GGC	GAG	TTG	GGG	GAC	GAT	CTG	AGA	ATG	GGG	GAG	1288
K	A	G	E	L	R	S	T	T	G	K	A	G	386
AAA	GCA	GGT	GAG	CTT	CGG	AGC	ACG	ACT	GGG	AAG	GCA	GGT	1327
E	S	W	D	K	K	E	K	M	G	Q	P	Q	399
GAG	TCC	TGG	GAT	AAG	AAG	GAA	AAG	ATG	GGG	CAA	CCC	CAG	1366

FIG. 2 B

G	K	S	G	N	A	G	E	A	R	S	Q	T	412
GGT	AAG	TCC	GGG	AAC	GCA	GGT	GAA	GCT	CGG	AGT	CAG	ACA	1405
E	K	G	C	E	A	P	K	E	V	S	T	M	425
GAG	AAG	GGC	TGT	GAA	GCC	CCA	AAG	GAG	GTG	AGC	ACA	ATG	1444
V	E	S	P	A	A	P	G	K	G	G	W	P	438
GTG	GAG	TCG	CCA	GCA	GCT	CCT	GGG	AAG	GGA	GGC	TGG	CCA	1483
G	S	R	G	Q	E	A	E	E	P	C	S	R	451
GGA	AGC	CGT	GGG	CAG	GAA	GCA	GAG	GAG	CCC	TGC	TCA	AGA	1522
A	G	D	G	A	G	A	L	E	T	E	L	E	464
GCA	GGT	GAT	GGG	GCT	GGT	GCC	CTG	GAG	ACA	GAG	CTG	GAA	1561
G	P	S	Q	P	A	L	E	K	D	A	E	R	477
GGA	CCC	AGC	CAG	CCT	GCT	CTG	GAG	AAG	GAT	GCA	GAA	AGG	1600
P	R	I	R	K	E	N	Q	D	G	P	A	P	490
CCT	CGG	ATA	CGG	AAG	GAG	AAC	CAA	GAC	GGG	CCA	GCC	CCG	1639
Q	E	E	G	K	G	G	Q	S	R	D	S	D	503
CAG	GAG	GAG	GGC	AAA	GGA	GGC	CAG	AGC	AGA	GAC	TCA	GAC	1678
Q	A	P	E	D	R	W	Y	E	A	E	K	V	516
CAG	GCT	CCT	GAG	GAC	AGA	TGG	TAT	GAG	GCA	GAG	AAA	GTC	1717
W	L	A	Q	K	D	G	F	T	L	A	T	V	529
TGG	CTG	GCT	CAG	AAG	GAT	GGA	TTT	ACT	CTT	GCT	ACG	GTG	1756
L	K	P	D	E	G	T	A	D	L	P	A	G	542
CTA	AAG	CCA	GAT	GAG	GGA	ACA	GCA	GAC	CTG	CCA	GCA	GGA	1795
R	V	R	L	C	I	D	A	D	K	T	I	T	555
AGG	GTG	AGA	CTT	TGC	ATT	GAT	GCT	GAC	AAA	ACC	ATC	ACT	1834
E	V	D	E	E	H	V	H	R	A	N	P	P	568
GAG	GTG	GAT	GAG	GAG	CAT	GTC	CAT	CGG	GCC	AAC	CCT	CCT	1873
E	L	D	Q	V	E	D	L	A	S	L	I	S	581
GAG	CTG	GAC	CAG	GTC	GAG	GAC	CTG	GCC	TCT	CTC	ATC	AGT	1912
V	N	E	S	S	V	L	N	T	L	L	Q	R	594
GTC	AAC	GAA	TCC	AGT	GTC	CTG	AAC	ACG	CTT	CTG	CAG	CGC	1951
Y	K	A	Q	L	L	H	T	C	T	G	P	D	607
TAC	AAA	GCT	CAG	CTG	CTG	CAC	ACC	TGC	ACA	GGG	CCT	GAT	1990
L	I	V	L	Q	P	R	G	P	S	V	P	S	620
CTG	ATT	GTC	CTC	CAG	CCC	CGG	GGG	CCC	TCG	GTG	CCT	TCT	2029

FIG. 2c

A	G	K	V	P	K	G	R	R	D	G	L	P	633
GCA	GGG	AAG	GTG	CCC	AAG	GGC	CGC	CGG	GAT	GGC	CTG	CCT	2068
A	H	I	G	S	M	A	Q	R	A	Y	W	A	646
GCC	CAC	ATT	GGC	TCC	ATG	GCA	CAG	CGG	GCA	TAC	TGG	GCG	2107
L	L	N	Q	R	R	D	Q	S	I	V	A	L	659
CTG	CTG	AAC	CAG	CGG	AGA	GAC	CAG	AGC	ATT	GTG	GCC	CTG	2146
G	R	S	G	A	G	K	T	T	C	C	E	Q	672
GGC	CGG	AGT	GGC	GCT	GGG	AAG	ACC	ACC	TGC	TGT	GAG	CAG	2185
V	L	E	H	L	V	G	M	A	G	S	V	D	685
GTC	CTG	GAA	CAC	CTG	GTG	GGG	ATG	GCA	GGC	AGT	GTG	GAT	2224
G	R	V	S	V	E	K	I	R	A	T	F	T	698
GGC	AGG	GTC	TCA	GTG	GAG	AAG	ATC	CGA	GCC	ACC	TTC	ACT	2263
V	L	R	A	F	G	S	V	S	M	A	H	S	711
GTC	CTC	CGG	GCC	TTC	GGC	TCT	GTG	TCC	ATG	GCC	CAC	AGC	2302
R	S	A	T	R	F	S	M	V	M	S	L	D	724
CGC	AGT	GCC	ACC	CGG	TTC	TCC	ATG	GTG	ATG	TCG	CTG	GAC	2341
F	N	A	T	G	R	I	T	A	A	Q	L	Q	737
TTC	AAC	GCT	ACA	GGC	CGC	ATC	ACA	GCT	GCT	CAG	CTC	CAG	2380
T	M	L	L	E	K	S	R	V	A	R	Q	P	750
ACA	ATG	CTT	TTG	GAG	AAG	AGC	CGC	GTG	GCA	CGG	CAG	CCG	2419
E	G	E	S	N	F	L	V	F	S	Q	M	L	763
GAA	GGG	GAA	AGT	AAC	TTC	CTG	GTT	TTC	TCC	CAG	ATG	CTG	2458
A	G	L	D	L	D	L	R	T	E	L	N	L	776
GCT	GGA	TTG	GAC	TTG	GAT	CTC	AGG	ACG	GAG	CTG	AAC	CTG	2497
H	Q	M	A	D	S	S	S	F	G	M	G	V	789
CAC	CAG	ATG	GCA	GAT	AGC	AGC	TCC	TTT	GGC	ATG	GGC	GTG	2536
W	S	K	P	E	D	K	Q	K	A	A	A	A	802
TGG	TCC	AAG	CCT	GAA	GAT	AAA	CAG	AAG	GCG	GCA	GCT	GCC	2575
F	A	Q	L	Q	G	A	M	E	M	L	G	I	815
TTT	GCC	CAG	CTC	CAG	GGT	GCC	ATG	GAG	ATG	CTG	GGC	ATC	2614
S	E	S	E	Q	R	A	V	W	R	V	L	A	828
TCA	GAG	AGC	GAG	CAG	CGG	GCT	GTT	TGG	CGG	GTC	CTG	GCA	2653
A	I	Y	H	L	G	A	A	G	A	C	K	V	841
GCC	ATC	TAC	CAC	CTG	GGC	GCG	GCG	GGG	GCC	TGC	AAA	GTG	2692
G	R	K	Q	F	M	R	F	E	W	A	N	Y	854

FIG. 2D

GGT	CGG	AAG	CAG	TTC	ATG	AGG	TTT	GAG	TGG	GCA	AAC	TAC	2731
A	A	E	A	L	G	C	E	Y	E	E	L	N	867
GCA	GCT	GAG	GCC	CTG	GGC	TGC	GAG	TAT	GAG	GAG	CTG	AAC	2770
T	A	T	F	K	H	H	L	R	Q	I	I	Q	880
ACG	GCC	ACC	TTC	AAG	CAC	CAC	CTT	CGA	CAG	ATC	ATC	CAG	2809
Q	M	T	F	G	P	S	R	W	G	L	E	D	893
CAA	ATG	ACG	TTT	GGG	CCA	AGC	CGA	TGG	GGC	CTC	GAG	GAT	2848
E	E	T	S	S	G	L	K	M	T	G	V	D	906
GAG	GAA	ACC	AGC	TCA	GGG	CTC	AAG	ATG	ACA	GGA	GTG	GAC	2887
C	V	E	G	M	A	S	G	L	Y	Q	E	L	919
TGT	GTG	GAG	GGG	ATG	GCC	TCG	GGC	CTG	TAC	CAG	GAA	CTC	2926
F	A	A	V	V	S	L	I	N	R	S	F	S	932
TTT	GCG	GCT	GTG	GTC	TCA	CTC	ATC	AAC	AGA	TCC	TTT	TCC	2965
S	H	H	L	S	M	A	S	I	M	V	V	D	945
TCC	CAC	CAT	CTC	TCC	ATG	GCC	TCC	ATC	ATG	GTG	GTG	GAC	3004
S	P	G	F	Q	N	P	R	H	Q	G	K	D	958
TCT	CCA	GGC	TTC	CAG	AAC	CCC	CGG	CAC	CAG	GGC	AAG	GAC	3043
R	A	A	T	F	E	E	L	C	H	N	Y	A	971
CGG	GCG	GCC	ACC	TTT	GAG	GAG	CTG	TGC	CAC	AAC	TAC	GCC	3082
H	E	R	L	Q	L	L	F	Y	Q	R	T	F	984
CAT	GAG	CGC	CTG	CAG	CTG	CTG	TTC	TAC	CAG	CGG	ACC	TTT	3121
V	S	T	L	Q	R	Y	Q	E	E	G	V	P	997
GTC	TCC	ACG	CTA	CAG	CGA	TAT	CAA	GAG	GAA	GGT	GTT	CCT	3160
V	Q	F	D	L	P	D	P	S	P	G	T	T	1010
GTG	CAG	TTT	GAC	CTC	CCG	GAC	CCC	TCC	CCA	GGG	ACC	ACC	3199
V	A	V	V	D	Q	N	P	S	Q	Q	V	R	1023
GTG	GCT	GTT	GTG	GAT	CAA	AAT	CCC	TCT	CAG	CAG	GTC	CGC	3238
L	P	A	G	G	G	A	Q	D	A	R	G	L	1036
TTA	CCA	GCT	GGA	GGA	GGT	GCC	CAG	GAT	GCC	AGA	GGC	CTT	3277
F	W	V	L	D	E	E	V	H	V	E	G	S	1049
TTC	TGG	GTC	TTA	GAT	GAG	GAA	GTC	CAT	GTA	GAG	GGC	TCC	3316
S	D	S	V	V	L	E	R	L	C	A	T	F	1062
AGT	GAC	AGT	GTG	GTG	CTC	GAG	CGT	CTG	TGT	GCT	ACT	TTC	3355
E	K	K	G	A	G	T	E	G	S	S	A	L	1075
GAG	AAG	AAA	GGA	GCT	GGG	ACT	GAA	GGG	TCC	TCT	GCC	CTG	3394

FIG. 2E

R	T	C	E	Q	P	L	Q	C	E	I	F	H	1088
CGG	ACC	TGT	GAG	CAG	CCC	CTC	CAG	TGT	GAG	ATT	TTC	CAC	3433
Q	L	G	W	D	P	V	R	Y	D	L	T	G	1101
CAG	TTG	GGA	TGG	GAC	CCT	GTG	CGG	TAC	GAC	CTC	ACG	GGC	3472
W	L	H	R	A	K	P	N	L	S	A	L	D	1114
TGG	CTC	CAC	AGA	GCC	AAG	CCC	AAC	CTC	TCG	GCC	CTG	GAT	3511
A	P	Q	V	L	Q	Q	S	K	R	E	E	L	1127
GCA	CCC	CAG	GTC	CTG	CAA	CAG	TCA	AAA	AGA	GAG	GAG	CTG	3550
R	S	L	F	Q	A	R	A	K	L	P	P	V	1140
CGG	AGT	CTA	TTC	CAG	GCC	CGG	GCC	AAG	CTG	CCT	CCT	GTG	3589
C	R	A	V	A	G	L	E	G	T	S	Q	Q	1153
TGC	CGG	GCT	GTG	GCA	GGC	CTG	GAG	GGC	ACC	TCC	CAG	CAG	3628
A	L	Q	R	S	R	M	V	R	R	T	F	A	1166
GCC	CTG	CAG	AGG	AGC	CGC	ATG	GTG	AGG	AGG	ACC	TTT	GCC	3667
S	S	L	A	A	V	R	R	K	A	P	C	S	1179
AGC	AGC	CTT	GCC	GCG	GTG	AGG	AGG	AAA	GCC	CCG	TGC	TCC	3706
Q	I	K	L	Q	M	D	A	L	T	S	M	I	1192
CAG	ATC	AAG	CTG	CAG	ATG	GAT	GCG	CTG	ACC	AGC	ATG	ATC	3745
K	R	S	R	L	H	F	I	H	C	L	V	P	1205
AAA	AGG	TCC	CGG	CTG	CAC	TTT	ATC	CAC	TGC	CTG	GTA	CCA	3784
N	P	V	V	E	S	R	S	G	Q	E	S	P	1218
AAC	CCT	GTG	GTG	GAA	AGC	AGG	AGT	GGG	CAG	GAA	TCT	CCA	3823
P	P	P	Q	P	G	R	D	K	P	G	A	G	1231
CCA	CCA	CCG	CAG	CCT	GGT	AGA	GAC	AAG	CCT	GGG	GCA	GGT	3862
G	P	L	A	L	D	I	P	A	L	R	V	Q	1244
GGA	CCT	CTG	GCC	CTG	GAT	ATC	CCA	GCA	CTG	AGG	GTC	CAG	3901
L	A	G	F	H	I	L	E	A	L	R	L	H	1257
CTT	GCT	GGG	TTC	CAC	ATC	CTG	GAG	GCT	CTG	CGT	CTG	CAT	3940
R	T	G	Y	A	D	H	M	G	L	T	R	F	1270
AGG	ACA	GGC	TAT	GCT	GAC	CAC	ATG	GGG	CTC	ACT	CGC	TTC	3979
R	R	Q	F	Q	V	L	D	A	P	L	L	K	1283
CGC	CGG	CAA	TTC	CAG	GTG	CTG	GAC	GCT	CCA	CTC	CTG	AAG	4018
K	L	M	S	T	S	E	G	I	D	E	R	K	1296
AAG	CTC	ATG	TCG	ACC	TCC	GAG	GGA	ATA	GAT	GAA	AGG	AAG	4057

FIG. 2f

A	V	E	E	L	L	E	T	L	D	L	E	K	1309
GCC	GTG	GAG	GAG	CTC	CTG	GAG	ACC	CTG	GAT	CTG	GAA	AAG	4096
K	A	V	A	V	G	H	S	Q	V	F	L	K	1322
AAG	GCG	GTG	GCT	GTG	GGG	CAC	AGC	CAA	GTT	TTT	CTC	AAG	4135
A	G	V	I	S	R	L	E	K	Q	R	E	K	1335
GCA	GGT	GTG	ATC	TCC	AGG	CTT	GAG	AAG	CAG	CGA	GAG	AAG	4174
L	V	S	Q	S	I	V	L	F	Q	A	A	C	1348
CTG	GTA	TCT	CAG	AGC	ATC	GTT	CTC	TTC	CAG	GCG	GCT	TGC	4213
K	G	F	L	S	R	Q	E	F	K	K	L	K	1361
AAG	GGC	TTT	CTG	TCT	CGC	CAG	GAA	TTC	AAG	AAG	CTG	AAG	4252
I	R	R	L	A	A	Q	C	I	Q	K	N	V	1374
ATT	CGC	CGA	CTG	GCT	GCA	CAG	TGC	ATC	CAG	AAG	AAT	GTG	4291
A	V	F	L	A	V	K	D	W	P	W	W	Q	1387
GCT	GTG	TTC	CTC	GCA	GTC	AAG	GAC	TGG	CCA	TGG	TGG	CAG	4330
L	L	G	S	L	Q	P	L	L	S	A	T	I	1400
CTG	CTT	GGT	TCC	CTC	CAG	CCT	CTA	CTT	AGT	GCC	ACC	ATT	4369
G	T	E	Q	L	R	A	K	E	E	E	L	T	1413
GGA	ACT	GAG	CAG	CTC	CGA	GCC	AAG	GAG	GAG	GAG	CTT	ACA	4408
T	L	R	R	K	L	E	K	S	E	K	L	R	1426
ACG	CTA	AGA	CGG	AAG	CTA	GAA	AAA	TCA	GAG	AAG	TTG	CGG	4447
N	E	L	R	Q	N	T	D	L	L	E	S	K	1439
AAT	GAA	CTC	CGG	CAG	AAC	ACA	GAT	CTG	CTA	GAA	AGC	AAG	4486
I	A	D	L	T	S	D	L	A	D	E	R	F	1452
ATT	GCT	GAC	TTG	ACC	TCT	GAC	CTT	GCC	GAT	GAG	CGC	TTC	4525
K	G	D	V	A	C	Q	V	L	E	S	E	R	1465
AAA	GGT	GAT	GTG	GCC	TGC	CAG	GTG	CTG	GAG	AGT	GAG	CGG	4564
A	E	R	L	Q	A	F	R	E	V	Q	E	L	1478
GCA	GAG	CGG	CTA	CAG	GCC	TTC	CGG	GAG	GTC	CAG	GAG	CTC	4603
K	S	K	H	E	Q	V	Q	K	K	L	G	D	1491
AAG	AGC	AAG	CAT	GAA	CAA	GTC	CAG	AAA	AAA	CTG	GGA	GAT	4642
V	N	K	Q	L	E	E	A	Q	Q	K	I	Q	1504
GTG	AAT	AAA	CAG	TTG	GAA	GAA	GCC	CAG	CAG	AAA	ATT	CAG	4681
L	N	D	L	E	R	D	P	T	G	G	A	D	1517
TTG	AAT	GAC	TTG	GAA	AGG	GAT	CCC	ACT	GGA	GGA	GCA	GAC	4720
E	W	Q	M	R	F	D	C	A	Q	M	E	N	1530

FIG. 26

GAG	TGG	CAG	ATG	CGC	TTC	GAC	TGT	GCT	CAG	ATG	GAG	AAC	4759
E	F	L	R	K	R	L	Q	Q	C	E	E	R	1543
GAG	TTC	CTC	AGA	AAG	CGT	CTG	CAG	CAA	TGC	GAG	GAG	AGG	4798
L	D	S	E	L	T	A	R	K	E	L	E	Q	1556
CTG	GAC	TCG	GAG	CTG	ACA	GCC	AGG	AAA	GAG	CTG	GAG	CAA	4837
K	L	G	E	L	Q	S	A	Y	D	G	A	K	1569
AAG	CTT	GGG	GAG	TTG	CAA	AGT	GCT	TAT	GAC	GGG	GCC	AAG	4876
K	M	A	H	Q	L	K	R	K	C	H	H	L	1582
AAG	ATG	GCT	CAC	CAA	CTG	AAG	AGG	AAG	TGC	CAC	CAT	CTT	4915
T	C	D	L	E	D	T	C	V	L	L	E	N	1595
ACC	TGT	GAC	CTT	GAG	GAT	ACC	TGC	GTC	CTG	CTA	GAG	AAC	4954
Q	Q	S	R	N	H	E	L	E	K	K	Q	K	1608
CAA	CAA	AGT	CGA	AAC	CAT	GAG	CTG	GAG	AAG	AAG	CAG	AAG	4993
K	F	D	L	Q	L	A	Q	A	L	G	E	S	1621
AAG	TTT	GAC	CTG	CAG	CTG	GCC	CAG	GCC	CTA	GGT	GAG	TCA	5032
V	F	E	K	G	L	R	E	K	V	T	Q	E	1634
GTG	TTT	GAG	AAG	GGT	CTC	CGT	GAG	AAA	GTG	ACC	CAG	GAG	5071
N	T	S	V	R	W	E	L	G	Q	L	Q	Q	1647
AAC	ACC	AGT	GTC	CGG	TGG	GAG	CTA	GGC	CAG	CTT	CAG	CAG	5110
Q	L	K	Q	K	E	Q	E	A	S	Q	L	K	1660
CAG	CTG	AAG	CAA	AAG	GAG	CAG	GAA	GCC	TCA	CAG	CTG	AAG	5149
Q	Q	V	E	M	L	Q	D	H	K	R	E	L	1673
CAG	CAG	GTG	GAG	ATG	CTA	CAG	GAC	CAT	AAA	CGG	GAG	CTG	5188
L	G	S	P	S	L	G	E	N	C	V	A	G	1686
CTG	GGG	TCA	CCC	TCT	CTG	GGG	GAA	AAT	TGC	GTT	GCT	GGC	5227
L	K	E	R	L	W	K	L	E	S	S	A	L	1699
TTG	AAG	GAG	AGG	CTC	TGG	AAG	TTG	GAA	TCC	AGC	GCC	CTT	5266
E	Q	Q	K	I	Q	S	Q	Q	E	N	T	I	1712
GAG	CAA	CAG	AAA	ATC	CAG	AGC	CAG	CAG	GAA	AAC	ACC	ATC	5305
K	Q	L	E	Q	L	R	Q	R	F	E	L	E	1725
AAG	CAG	CTG	GAG	CAG	CTC	CGC	CAG	CGG	TTT	GAG	CTG	GAG	5344
I	E	R	M	K	Q	M	H	Q	K	D	R	E	1738
ATC	GAG	CGG	ATG	AAG	CAG	ATG	CAC	CAG	AAG	GAC	CGT	GAG	5383
D	Q	E	E	E	L	E	D	V	R	Q	S	C	1751
GAC	CAG	GAG	GAG	GAA	CTG	GAG	GAT	GTC	CGT	CAG	TCC	TGC	5422

FIG. 2H

Q	K	R	L	H	Q	L	E	M	Q	L	E	Q	1764
CAG	AAG	CGG	CTT	CAT	CAG	CTG	GAA	ATG	CAG	CTG	GAG	CAA	5461
E	Y	E	E	K	Q	M	V	L	H	E	K	Q	1777
GAG	TAT	GAA	GAG	AAG	CAG	ATG	GTC	CTC	CAT	GAG	AAG	CAA	5500
D	L	E	G	L	I	G	T	L	C	D	Q	I	1790
GAT	TTG	GAA	GGC	TTG	ATC	GGA	ACC	CTC	TGT	GAC	CAG	ATT	5539
G	H	R	D	F	D	V	E	K	R	L	R	R	1803
GGC	CAT	CGG	GAC	TTT	GAT	GTG	GAG	AAG	CGA	CTT	CGG	AGA	5578
D	L	R	R	T	H	A	L	L	S	D	V	Q	1816
GAC	CTC	AGG	AGG	ACA	CAT	GCA	CTG	TTG	TCA	GAC	GTG	CAG	5617
L	L	L	G	T	M	E	D	G	K	T	S	V	1829
CTC	CTT	CTG	GGC	ACC	ATG	GAG	GAT	GGC	AAG	ACA	TCA	GTC	5656
S	K	E	E	L	E	K	V	H	S	Q	L	E	1842
AGC	AAG	GAG	GAG	CTG	GAG	AAA	GTG	CAC	AGC	CAG	CTG	GAG	5695
Q	S	E	A	K	C	E	E	A	L	K	T	Q	1855
CAG	AGT	GAA	GCC	AAG	TGT	GAG	GAG	GCC	TTG	AAG	ACG	CAG	5734
K	V	L	T	A	D	L	E	S	M	H	S	E	1868
AAG	GTG	CTC	ACA	GCG	GAC	CTG	GAG	AGC	ATG	CAC	AGC	GAG	5773
L	E	N	M	T	R	N	K	S	L	V	D	E	1881
CTG	GAG	AAC	ATG	ACG	CGG	AAC	AAG	AGC	CTG	GTG	GAT	GAG	5812
Q	L	Y	R	L	Q	F	E	K	A	D	L	L	1894
CAG	CTG	TAC	AGG	CTG	CAG	TTT	GAG	AAG	GCG	GAC	CTC	CTG	5851
K	R	I	D	E	D	Q	D	D	L	N	E	L	1907
AAG	CGC	ATC	GAT	GAG	GAC	CAG	GAT	GAC	CTG	AAT	GAG	CTG	5890
M	Q	K	H	K	D	L	I	A	Q	S	A	A	1920
ATG	CAG	AAG	CAC	AAG	GAC	CTC	ATT	GCT	CAG	TCT	GCT	GCT	5929
D	I	G	Q	I	Q	E	L	Q	L	Q	L	E	1933
GAC	ATT	GGG	CAG	ATC	CAA	GAA	CTG	CAG	CTG	CAG	CTG	GAG	5968
E	A	K	K	E	K	H	K	L	Q	E	Q	L	1946
GAA	GCC	AAG	AAG	GAG	AAG	CAC	AAG	CTA	CAA	GAA	CAA	CTG	6007
Q	V	A	Q	M	R	I	E	Y	L	E	Q	S	1959
CAG	GTG	GCT	CAG	ATG	CGC	ATC	GAG	TAC	CTG	GAA	CAG	TCC	6046
T	V	D	R	A	I	V	S	R	Q	E	A	V	1972
ACC	GTG	GAT	CGA	GCC	ATC	GTC	AGC	AGG	CAG	GAG	GCG	GTC	6085

FIG. 2I

I	C	D	L	E	N	K	T	E	F	Q	K	V	1985
ATC	TGT	GAC	CTA	GAG	AAC	AAG	ACA	GAG	TTC	CAG	AAG	GTG	6124
Q	I	K	R	F	E	V	L	V	I	R	L	R	1998
CAG	ATT	AAG	AGA	TTT	GAG	GTC	CTG	GTG	ATC	CGG	CTT	CGG	6163
D	S	L	I	K	M	G	E	E	L	S	Q	A	2011
GAC	AGC	CTG	ATC	AAG	ATG	GGG	GAG	GAG	CTT	TCA	CAG	GCG	6202
A	T	S	E	S	Q	Q	R	E	S	S	Q	Y	2024
GCC	ACC	TCC	GAG	TCC	CAG	CAG	CGG	GAG	AGC	AGC	CAG	TAC	6241
Y	Q	R	R	L	E	E	L	K	A	D	M	E	2037
TAC	CAG	CGG	CGC	CTG	GAA	GAG	CTG	AAG	GCC	GAC	ATG	GAA	6280
E	L	V	Q	R	E	A	E	A	S	R	R	C	2050
GAG	CTG	GTG	CAG	CGG	GAG	GCA	GAG	GCC	AGC	CGG	CGG	TGC	6319
M	E	L	E	K	Y	V	E	E	L	A	A	V	2063
ATG	GAG	CTG	GAG	AAG	TAC	GTG	GAG	GAA	CTT	GCA	GCA	GTG	6358
R	Q	T	L	Q	T	D	L	E	T	S	I	R	2076
AGG	CAA	ACC	CTC	CAG	ACA	GAC	CTG	GAG	ACA	TCC	ATT	CGG	6397
R	I	A	D	L	Q	A	A	L	E	E	V	A	2089
CGG	ATT	GCC	GAC	CTG	CAG	GCT	GCC	TTG	GAA	GAA	GTG	GCA	6436
S	S	D	S	D	T	E	S	V	Q	T	A	V	2102
TCC	AGT	GAC	AGT	GAT	ACT	GAG	AGT	GTC	CAG	ACG	GCA	GTG	6475
D	C	G	S	S	G	R	K	E	M	D	N	V	2115
GAT	TGT	GGC	AGC	AGC	GGC	CGA	AAA	GAG	ATG	GAT	AAC	GTC	6514
S	I	L	S	S	Q	P	E	G	S	L	Q	S	2128
TCC	ATC	CTC	AGC	TCC	CAG	CCA	GAG	GGC	AGC	CTG	CAG	TCC	6553
W	L	S	C	T	L	S	L	A	T	D	T	M	2141
TGG	TTG	AGC	TGT	ACT	CTG	TCC	CTG	GCC	ACA	GAT	ACT	ATG	6592
R	T	P	S	R	Q	S	A	T	S	S	R	I	2154
AGG	ACT	CCT	TCT	CGA	CAG	TCA	GCC	ACC	AGC	AGC	CGC	ATC	6631
L	S	P	R	I	N	E	E	A	G	D	T	E	2167
CTC	AGC	CCC	AGG	ATA	AAC	GAA	GAG	GCT	GGG	GAC	ACT	GAG	6670
R	T	Q	S	A	L	A	L	S	R	A	R	S	2180
AGG	ACC	CAG	TCG	GCA	TTG	GCA	CTG	AGC	AGA	GCC	CGG	TCC	6709
T	N	V	H	S	K	T	S	G	D	K	P	V	2193
ACC	AAT	GTC	CAC	AGC	AAG	ACC	TCA	GGA	GAC	AAG	CCT	GTT	6748
S	P	H	F	V	R	R	Q	K	Y	C	H	F	2206

FIG. 2J

TCT	CCC	CAC	TTT	GTC	CGC	CGG	CAA	AAG	TAC	TGT	CAT	TTT	6787
G	D	G	E	V	L	A	V	Q	R	K	S	T	2219
GGG	GAC	GGC	GAA	GTG	CTT	GCC	GTC	CAG	AGA	AAG	TCC	ACA	6826
E	R	L	E	P	A	S	S	P	L	A	S	R	2232
GAG	AGA	TTA	GAA	CCT	GCT	TCC	TCT	CCC	CTG	GCT	TCT	CGG	6865
S	T	N	T	S	P	L	S	R	E	K	L	P	2245
AGT	ACA	AAT	ACA	TCC	CCG	CTG	TCG	AGG	GAA	AAG	CTG	CCC	6904
S	P	S	A	A	L	S	E	F	V	E	G	L	2258
AGT	CCT	TCA	GCG	GCC	CTC	TCG	GAG	TTC	GTG	GAA	GGG	CTC	6943
R	R	K	R	A	Q	R	G	Q	G	S	T	L	2271
CGG	AGG	AAG	AGA	GCC	CAG	AGA	GGC	CAG	GGG	TCC	ACG	CTG	6982
G	L	E	D	W	P	T	L	P	I	Y	Q	T	2284
GGC	CTA	GAG	GAC	TGG	CCC	ACT	CTC	CCC	ATT	TAC	CAG	ACG	7021
T	G	A	S	T	L	R	R	G	R	A	G	S	2297
ACT	GGG	GCC	TCC	ACA	CTA	AGG	AGG	GGC	AGG	GCT	GGC	AGT	7060
D	E	G	N	L	S	L	R	V	G	A	K	S	2310
GAC	GAG	GGA	AAC	CTC	TCG	CTG	AGG	GTT	GGG	GCA	AAG	TCA	7099
P	L	E	I	E	G	A	A	G	G	L	L	R	2323
CCC	CTG	GAA	ATC	GAA	GGG	GCC	GCT	GGT	GGT	CTC	TTG	AGG	7138
S	T	S	L	K	C	I	S	S	D	G	V	G	2336
TCC	ACC	AGC	CTC	AAA	TGC	ATC	TCT	TCA	GAC	GGT	GTT	GGG	7177
G	T	T	L	L	P	E	K	S	K	T	Q	F	2349
GGC	ACA	ACC	CTA	CTC	CCC	GAA	AAG	TCG	AAA	ACC	CAA	TTC	7216
S	S	C	E	S	L	L	E	S	R	P	S	M	2362
AGT	TCC	TGC	GAG	TCC	CTC	TTA	GAA	TCC	AGA	CCG	AGC	ATG	7255
G	R	K	L	S	S	P	T	T	P	R	D	M	2375
GGG	AGA	AAA	CTG	AGC	TCT	CCG	ACC	ACA	CCC	AGG	GAC	ATG	7294
L	L	S	P	T	L	R	P	R	R	R	C	L	2388
CTG	TTG	TCG	CCC	ACA	CTG	CGT	CCT	CGG	AGG	CGG	TGT	CTG	7333
E	S	S	V	D	D	A	G	C	P	D	L	G	2401
GAG	TCC	TCT	GTG	GAC	GAT	GCG	GGC	TGT	CCA	GAC	CTT	GGA	7372
K	E	P	L	V	F	Q	N	R	Q	F	A	H	2414
AAG	GAG	CCG	CTT	GTT	TTC	CAG	AAC	CGC	CAG	TTT	GCC	CAC	7411
L	M	E	E	P	L	G	S	D	P	F	S	W	2427
CTG	ATG	GAG	GAA	CCT	CTA	GGC	AGT	GAC	CCA	TTC	AGC	TGG	7450

FIG. 2K

K	L	P	S	L	D	Y	E	R	K	T	K	V	2440
AAA	CTC	CCA	AGC	CTC	GAC	TAC	GAA	CGC	AAG	ACC	AAA	GTG	7489
D	F	D	D	F	L	P	A	I	R	K	P	Q	2453
GAC	TTC	GAT	GAC	TTC	CTC	CCA	GCT	ATC	CGG	AAG	CCC	CAG	7528
T	P	T	S	L	A	G	S	A	K	G	G	Q	2466
ACA	CCT	ACC	TCC	TTG	GCT	GGA	TCA	GCC	AAA	GGT	GGG	CAA	7567
D	G	S	Q	R	S	S	I	H	F	E	T	E	2479
GAC	GGT	TCA	CAG	CGT	TCA	AGC	ATC	CAC	TTT	GAA	ACG	GAA	7606
E	A	N	R	S	F	L	S	G	I	K	T	I	2492
GAG	GCT	AAC	CGT	TCC	TTT	CTC	TCG	GGG	ATC	AAG	ACC	ATT	7645
L	K	K	S	P	E	P	K	E	D	P	A	H	2505
TTG	AAG	AAG	AGC	CCG	GAG	CCC	AAG	GAG	GAT	CCC	GCT	CAC	7684
L	S	D	S	S	S	S	S	G	S	I	V	S	2518
CTG	TCT	GAC	TCG	TCC	TCA	TCC	TCC	GGC	TCC	ATC	GTG	TCC	7723
F	K	S	A	D	S	I	K	S	R	P	G	I	2531
TTC	AAA	AGT	GCT	GAC	AGC	ATC	AAA	AGT	CGA	CCA	GGA	ATC	7762
P	R	L	A	G	D	G	G	E	R	T	S	P	2544
CCA	CGA	CTT	GCG	GGT	GAC	GGT	GGC	GAG	CGA	ACG	TCC	CCC	7801
E	R	R	E	P	G	T	G	R	K	D	D	D	2557
GAG	CGG	AGA	GAG	CCA	GGG	ACG	GGG	AGG	AAA	GAC	GAC	GAT	7840
V	A	S	I	M	K	K	Y	L	Q	K	*		2569
GTT	GCG	AGC	ATA	ATG	AAG	AAA	TAC	CTC	CAG	AAG	TAG	gaa	7879
ccagttcagcctccttgaagctgcccttgaagacttcccgactctacaataa												7931	
cttgagagacagagagactggccaggcctccccggtggccagagccagccagc												7983	
atggccaccctcaagaggcgagatgagcccacagaggcatatcctgcgggga												8035	
tgctgggctcccagtggtgttgctgaacaaaataaagtgttgactcctgg												8087	
gcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa												8117	

FIG. 2L

HGDMLP-1	APEDRWYEAETKVVLAQKDGFTLVLKPDDEGTADLPAGRVRLCIDADTKTITEVDEEHVHR	60
BAA93660	AAEEAWYETEKVWLVHRDGFSLASQLKSEE--LSLPEGARVKLDHGDGAILDVEDDIEK	58
BAA13206	-----	
HGDMLP-1	ANPELQVEDLASLISVNESSVLNTLLQRYKAQLLHTCTGPDLLVLPQRP-----SVPS	116
BAA93660	ANAPSCDRLEDLASLVYLNESVLTLLRQRYGASLLHTYAGPSLLVLSRGPAPVYSEKV	118
BAA13206	-----	
HGDMLP-1	AGKVPKGRRDGLPAHIGSMAQRAYWALLNQRRDQSI VALGRSGAGKTTCCCEQVLEHLVGM	176
BAA93660	MHMFKGCRRREDMAPIH YAVAQTAYRAMLSRQDQSI VLLGSSGSGKTTSFQHLVQYLATI	178
BAA13206	MHMFKGCRRREDMAPIH YAVAQTAYRAMLSRQDQSI ILLGSSGSGKTTSCQHLVQYLATI	60
	. **:::..** ::** ** ** * .*:****: ** ***:****. :::::..* .:	
HGDMLP-1	AGSVDGRVS-VEKIRATFTVLRAFGSVSMAHSRSATRFMSVMSLDFNATGRITAAQLQTM	235
BAA93660	AGTSGTKVFSVEKWQALSTLLEAFGNSPTIMNGSATRFSQLSLDFDQAGQVASASIQTM	238
BAA13206	AGISGNKVFSEKWQALYTLLEAFGNSPTIINGNATRFSQLSLDFDQAGQVASASIQTM	120
	** . : * *** : * * * : * . * . * . * . * . * . * . * . * . * . * . * . *	
HGDMLP-1	LLEKSRVARQPEGESNFLVFSQMLAGLDLRLTELNLHQADSSSFGMGVWSKPEDKQKA	295
BAA93660	LLEKLRVARRPASEATFNVFYLLACGDATLRLTELHLNHLAENNVFGIVPLSKPEEKQKA	298
BAA13206	LLEKLRVARRPASEATFNVFYLLACGDGTLRLTELHLNHLAENNVFGIVPLAKPEEKQKA	180
	**** ***: * .*: * * * : * * * * : * . * . * . * . * . * . * . * . *	
HGDMLP-1	AAFAQLQGAMEMLGISESEQRVWVRLAAIYHLGAAG---ACKVGRKQFMREFEWANYA	351
BAA93660	AQQFSKLQAAMKVLAI SPEEQKTCWLILASIYHLGAAGATKEAAEAGRKQFARHEWAQKA	358
BAA13206	AQQFSKLQAAMKVLGISPDEQKACWFILAAIYHLGAAGATKEAAEAGRKQFARHEWAQKA	240
	* *: * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . *	
HGDMLP-1	AEALGCEYEELNTATFKHHLRQIIQQMTFGPSRWGLEDEETSSGL--KMTGVDCVEGMAS	409
BAA93660	AYLLGCSLEELSSAIFKHQLKGGTLQRSTS-FRQGPESGLGEGT--KLSALECLEGMAS	415
BAA13206	AYLLGCSLEELSSAIFKHQHKGGTLQRSTS-FRQGPESGLGDGTGPKLSALECLEGMAA	299
	* ***. ***.: * ***: : * : . * * * : . . * * * : : : : : * : * : *	
HGDMLP-1	GLYQELFAAVVSLINRSFSSHLSMASIMVVDSPGFQNPQRHOGKDRAATFEELCHNYAHE	469
BAA93660	GLYSELFTLLISLVNRALKSSQHSLSMMIVDTPGFQNPQEWGGSARGASFEELCHNYAQD	475
BAA13206	GLYSELFTLLVSLVNRALKSSQHSLSMMIVDTPGFQNPQEGGSARGASFEELCHNYTQD	359
	.: : ***:***: * : * .*:***:*****. * . * .*:***:***: : :	
HGDMLP-1	RLQLLFYQRTFVSTLQRYQEEGVPVQFDLPDPSPGTTVAVVDQNPSSQVRLPAGGGAQDA	529
BAA93660	RLQRLFHERTFLOELERYKEDNIELAFDDLEPVADDSVAAVDQ-ASH--LVRSLAHADDEA	532
BAA13206	RLQRLFHERTFVQELERYKEENIELAFDDLEPPTDDSVAAVDQ-ASHQSLVRSRLARTDEA	418
	*** ***:***. * :***:..: : * * * : * . . * . * . * . * . * . * . *	
HGDMLP-1	RGLFWVLDEEVHVEGSSDSVVLRLCATFEKKGAGTEGSSALRTCEQPLQCEIFHQLGWD	589
BAA93660	RGLLWLLEEEALVPGATEDALLDRLFSYYPQEGDKKGQSPLLRSSKPRHFLGHSHGTN	592
BAA13206	RGLLWLLEEEALVPGASEDTLLERLFSYYPQEGDKKGQSPLLHSSKPHHFLGHSHGTN	478
	::***. * * : : : : : : : : : : : . * . * . * . * . * . * . *	
HGDMLP-1	PVRYDLTGWLHRAKPNLSALDAPQVLQOSKREELRSLFQARAKLPVPCR-AVAGLEGTSQ	648
BAA93660	WVEYNVAGWLNLYTKQNPATQNPRLQDSQKKIISNLFGRAGSATVLSGSAGLEGSSQ	652
BAA13206	WVEYNVTGWLNYTKQNPATQNPRLQDSQKKIISNLFGRAGSATVLSGSAGLEGSSQ	538
	* .*:***: * * : : : : : : : : : . * . * . * . * . * . * . *	
HGDMLP-1	QALQSRMVRRFTFASSLAAVRRKAPCSQIKLQMDALTSMIKRSRLHFIHCLVPNPV---V	705
BAA93660	LALRRATSMRKTFTTGMAAVKKKSLCQIKLQVDALIDTIKRSKMHFVHCFLPVAEGWPG	712
BAA13206	LALRRATSMRKTFTTGMVAVKKKSLCQMKLQVDALIDTIKSKLHFVHCFLPVAEGWAG	598
	::***: * * : : : : : : : : : . * . * . * . * . * . * . *	

FIG. 3A

HGDMLP-1	ESRSQGESPPPPQPPGRDKPGAG---GPLALDI PALRVOLQAGFHIIEALRLHRTGYADHM	761
BAA93660	EPRSASSRRVSSSELDLPPGDPCEAGLLQLDVSLLRAQLRGSRLLDAMRMYRQGYPDHM	772
BAA13206	EPRSASSRRVSSSELDLPSGDHCEAGLLQLDVP LLRTQLRGSRLLDAMRMYRQGYPDHM	658
	*.***... .. * * .. * * * *: . ** . * : : : : : : * * . ***	
HGDMLP-1	GLTRFRRRQFQVLDAPLLKKLMSTSEGIDERKAVEELLETL DLEKKAVAVGHSQVFLKAGV	821
BAA93660	VFSEFRRRFDVLAPHLTKKHGRNYIVVDEKRAVEELLES DLEKSSCCLGLSRVFFRAGT	832
BAA13206	VFSEFRRRFDVLAPHLTKKHGRNYIVVDERRAVEELLECLDLEKSSCCMGLSRVFFRAGT	718
	: . : . *** : * : * . * * * . : : : : : ***** : : * * : : : : : *	
HGDMLP-1	ISRLEKQREKLVSQSIVLFQAACKGFLSRQEFKKLKIRRLAAQCIQKNVAVFLAVKDWPW	881
BAA93660	LARLEEQRDEQTSRHLTLFQAACRGYLARQHFKKRKIQDLAIRCVQKNIKKNKGVKDWPW	892
BAA13206	LARLEEQRDEQTSRNLTLFQAACRGYLAQHFKKRKIQDLAIRCVQKNIKKNKGVKDWPW	778
	: : *** : * : : . * : : . ***** : * : * : * : * : * : * : * : * : *	
HGDMLP-1	WQLLGSLOPILLSATIGTEQLRAKEEELTTLRRKLEKSEKLRNELRQNTD LLESKIADLTS	941
BAA93660	WKLFTTVRPLIQVQLSEEQIRNKDEEIQQLRSKLEKVEKERNELRLSSDRLETRISELTS	952
BAA13206	WKLFTTVRPLIEVQLSEEQIRNKDEEIQQLRSKLEKA EKERNELRLNSDRLESRISELTS	838
	* : : : : * : : . . * * * * * : * : * : * * * * * * * * * * : * * : : : : * : *	
HGDMLP-1	DLADERFKGDVACQVLESERAERLQAFREVQELKSKHEQVQKKLGDV NQLEEAQQKIQL	1001
BAA93660	ELTDERNTGESASQLLDAETAERLRTEKEMKELQTYDALKKQMEV MEMEVMEARLIRAA	1012
BAA13206	ELTDERNTGESASQLLDAETAERLRAEKEMKELQTYDALKKQMEV MEMEVMEARLIRAA	898
	: * : * * . * : * . : * : * : * * * : : : : : : : : : : : : : * : *	
HGDMLP-1	N-DLERDPTGGADEWQMRFDCAQMENEF LRKRLOQ-CEERL DSELTARKELEQKLGELOS	1059
BAA93660	EINGEVDDDDAGGEWRLKYERAVREVDFTKKRLOQEELEDKMEVEQQSRRLERRLGDLQA	1072
BAA13206	EINGEVDDDDAGGEWRLKYERAVREVDFTKKRLOQEFEDKLEVEQQNKRQLERRLGDLQA	958
	: : * * * : : : : * * : * : * * * * : : : : * : : : : : * : *	
HGDMLP-1	AYDGAKKMAHQ LKRKCHLTC DLEDTCV LLENQQSRNHELEKKQKKFDLQALQALGESVF	1119
BAA93660	DSDESQALQQLKKKCQR LTAELQDTKLHLEGQQVRNHELEKKQR RFDSLSQAHEETQR	1132
BAA13206	DSEESQALQQLKKKCQR LTAELQDTKLHLEGQQVRNHELEKKQR RFDSLSQAHEEAQR	1018
	: : : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
HGDMLP-1	EKGLREKVTQENTSVRWELGQLQQQLKQKEQEASQLKQQVEM LQDHKRELLGSPSLGENC	1179
BAA93660	EKLQREKLQREKMDLLAEAFSLKQQMEEKDLDIAGFTQKVVSLEAELQDISSQESKDEAS	1192
BAA13206	EKLQREKLQREKMDLLAEAFSLKQQLEEKMDIAGFTQKVVSLEAELQDISSQESKDEAS	1078
	** * : * : : * : * . : * : : : : : : : * : * * : . : : : . * * . *	
HGDMLP-1	VAGLKERLWKLESSALEQQKIQSQQENTIKQLEQLRQRFELIERM QMHQKDREDQEEE	1239
BAA93660	LAKVKKQLRDLEAKVKDQEEELDEQAGSIQMLEQAKLRLEMEMERMQTHSKEMESRDEE	1252
BAA13206	LAKVKKQLRDLEAKVKDQEEELDEQAGTIQMLEQAKLRLEMEMERMQTHSKEMESRDEE	1138
	: * : * : * . * : . . * : : . : * . : * : * * : * : * : * : * * : * : *	
HGDMLP-1	LEDVRQSCQKRLHQLEMQLEQEYEEKQMV LHEKQDLEGLIGTLCDQIGHRDFVEKRLRR	1299
BAA93660	VEEARQSCQKKLKQMEVQLEEEYEDKQKALREKRELESKLTSLSDQVNQRDFESEKRLRK	1312
BAA13206	VEEARQSCQKKLKQMEVQLEEEYEDKQKVLREKRELEGKLATLS DQVNRRDFESEKRLRK	1198
	: * : . * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
HGDMLP-1	DLRRTHALLSDVQ LLLGTMEDGCRVSKEELEKVHSQLEQSEACEEALKTQKVLTADLE	1359
BAA93660	DLKRTKALLADAQIMLDHLKN--NAPSKREIAQLKNQLEEESEFTCAA AVKARKAMEVEME	1370
BAA13206	DLKRTKALLADAQIMLDHLKN--SAPSKREIAQLKNQLEEESEFTCAA AVKARKAMEVEIE	1256
	* : * : * : * : * : * : * : : : : * : * : : : * : * : * : * : * : *	
HGDMLP-1	SMHSELENMTRNKSIVDEQYLRLQFEKADLLKRIDEQDDLNELMQKHKDLIAQSAADIG	1419
BAA93660	DLHLQIDDI AKATALEEQLSRLQREKNEIQNRLEEDQEDMELMKKHKAAVAQASRDMA	1430
BAA13206	DLHLQIDDI AKATALEEQLSRLQREKNEIQNRLEEDQEDMELMKKHKAAVAQASRDIA	1316
	. * : . : : : : * : : * : * * * : : : : : * : * : * : * : * : * : *	

FIG. 3B

HGDMLP-1	QIQELQLQLEEAKEKHEKHKLQEQLQVAQMRIEYLEQSTVDRAIVSRQEAVICDLENKTEFQ	1479
BAA93660	QMNDLQAQIEESNKEKQELQEKQLALQSQVEFLEQSMVDKSLVSRQEAKIRELETRLEFE	1490
BAA13206	QINDLQAQLEEANKKEKQELQEKQLALQSQVEFLEQSMVDKSLVSRQEAKIRELETRLEFE *:*** *::**::***::***:** * ::*:**** **::***** * :*.: **:	1376
HGDMLP-1	KVQIKRFEVLVIRLRDSLIMGEELSQAATSESQQRESSQYQRREELKADMEELVQRE	1539
BAA93660	KTQVKRLLENLASRLKETMEKLTEERDQRAAENREKEQNKRILQRLRDTKEEMSELARKE	1550
BAA13206	RTQVKRLESASRLKENMEKLTEERDQRIAAENREKEQNKRILQRLRDTKEEMGELARKE :.*:**:* *. **:.: : *: ** . * ::*:***:..: **:*.: * :* **.**:*	1436
HGDMLP-1	AEASRRCMELEKYVEELAARVQTLLTDLETISRRIADLQAALAEVASSDSDESQVTAVD	1599
BAA93660	AEASRKKHELEMDLESLEAANQSLQADLKLAFAKRGIDLQAIEDMESDENEDLINSEGD	1610
BAA13206	AEASRKKHELEMDLESLEAANQSLQADLKLAFAKRGIDLQAIEDMESDENEDLINSEGD *****: *** :*. * ..*:***:~::~**.******:~: .*:~: : ~: *	1496
HGDMLP-1	CGSSGRKEMDNVSILSSQPEGLSQSWLSCTLSLATDTMRTPSRQSATSRRILSPRINEEA	1659
BAA93660	S-----DVDSELEDVRDG-VKSWLS-----KNKGPSK-----	1636
BAA13206	S-----DVDSELEDVRDG-VKSWLS-----KNKGPSK----- . * * *..: * ::**** . : **:	1522
HGDMLP-1	GDTERTQSALASRARSTNVHSKTSGDKPVSHPFVRRQKYCHFGDGEVLAVORKSTERLE	1719
BAA93660	-----APSDDGSLKSSSPTSHWKPLAPDFSDE-----HDPVD	1669
BAA13206	-----AASDDGSLKSSSPTSYSWKS LAPDRSDE-----HDPLD * * * : * * ..:~* : ~:~	1555
HGDMLP-1	PASSPLASRSTNTSPLSREKLPSPSA	1745
BAA93660	SISRPFRFSHSYLSDSSTEAKLTETSA	1695
BAA13206	NTSRPRYSHSYLSDSSTEAKLTETNA * * *~* ... :. *~*.....	1581

FIG. 3 c

Sequences Aligned

Query: myocilin [Homo sapiens]: Length 504 (1 .. 504)

Sbjct: hGDMLP-1: Length 2568 (1 .. 2568)

Region of Optimal Alignment

```

Query: 82   RLDLEATKARLSSLESLLHQLTLDQAARPQETQEGLORELGTLRRERDQ-LETQ---TRE 137
           R DL  T A LS ++ LL  +   + + +E  E +  +L      + ++ L+TQ   T +
Sbjct: 1802 RDLRRTHALLSDVQLLLGTMEDGKTSVSKEELEKVHSQLEQSEAKCEEALKTQKVLTA 1861

Query: 138  LETAYS---NLLRDKSVLEEEKKRLRQENENLARRLESSSQEVARL 180
           LE+ +S   N+ R+KS+++E+  RL+ E  +L +R++   ++  L
Sbjct: 1862 LESMHSELENMTRNKSLVDEQLYRLQFEKADLLKRIDEDQDDLNEL 1907
  
```

Statistics

Identities = 30/106 (28%), Positives = 58/106 (54%), Gaps = 7/106 (6%)

FIG. 4

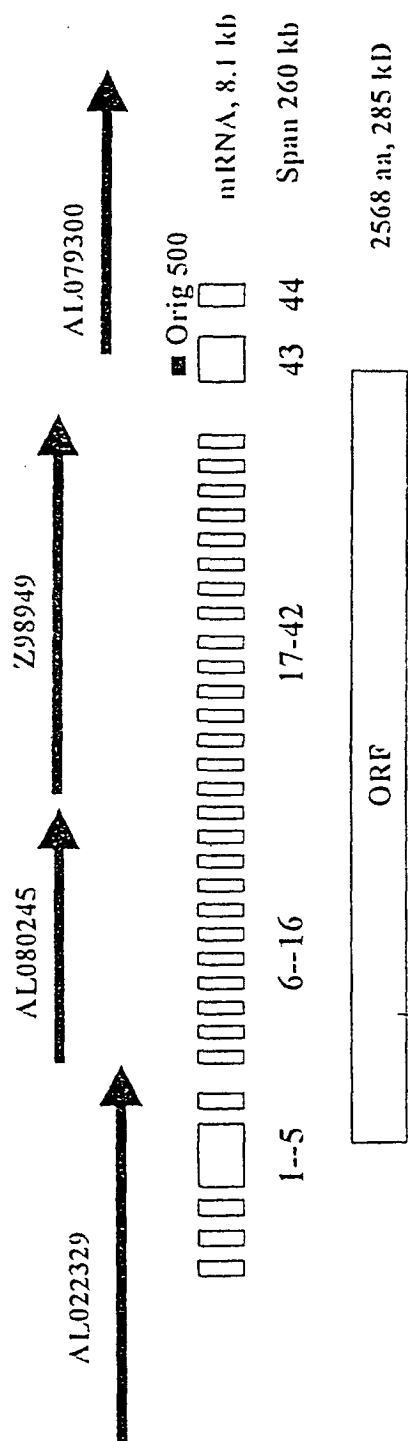
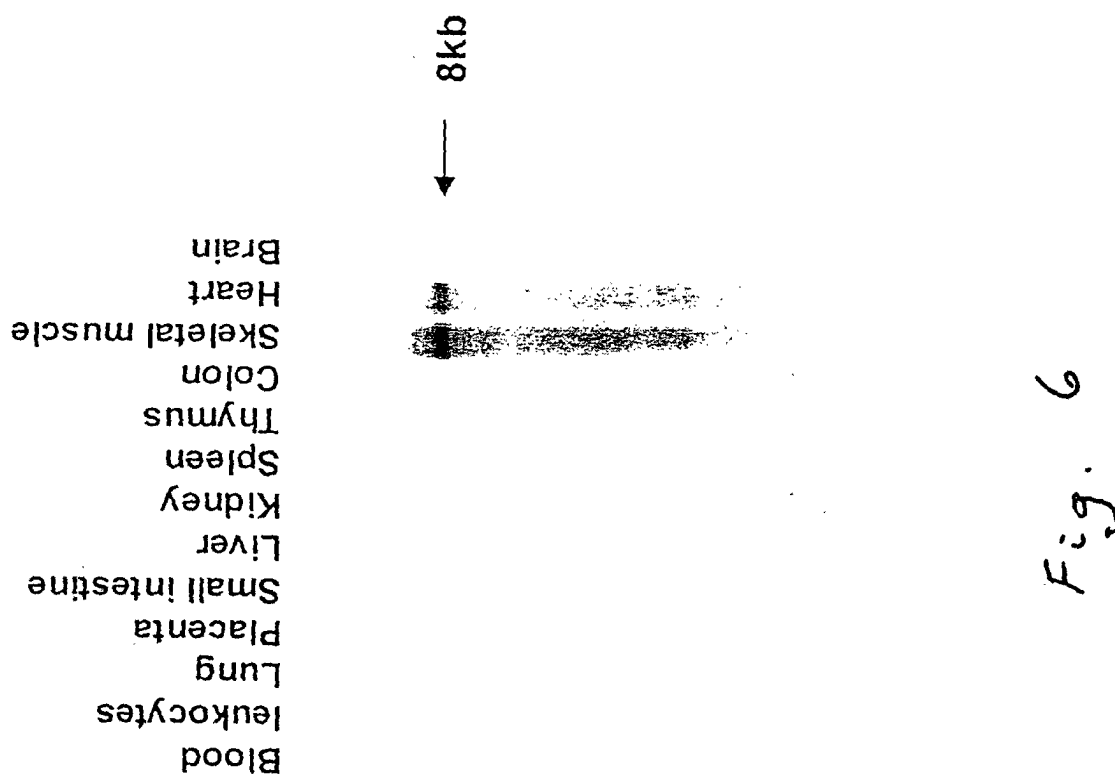


Fig. 5





European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

which under Rule 45 of the European Patent Convention EP 01 11 2637 shall be considered, for the purposes of subsequent proceedings, as the European search report

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	DATABASE EMBL [Online] Entry HS125H2, Acc.no. Z98949, 6 September 1997 (1997-09-06) LLOYD, D.: "Human DNA sequence from clone CTA-125H2 on chromosome 22q11-12 Contains part of the gene for a novel protein similar to KIAA0216 and myosin heavy chain, ESTs, GSSs and a CpG Island." XP002177437 see nt. 1-120630 ---	1-26	C12N15/12 C12N15/62 C07K14/47 C07K16/18 C12Q1/68 G01N33/53 A61K38/17 A61K31/7088
P,X	DATABASE EMBL [Online] Entry/Acc.no. AB042648, 16 May 2001 (2001-05-16) YOKOTA, J ET AL.: "Homo sapiens bk125H2.1 mRNA, complete cds." XP002177438 * the whole document * ---	1,4-10	
P,X	WO 00 58473 A (CURAGEN CORP ; LEACH MARTIN (US); SHIMKETS RICHARD A (US)) 5 October 2000 (2000-10-05) see seq.ID's 3195 and 3196 and paragraphs relating to ORF1598 --- -/--	1,4-21, 23-26	TECHNICAL FIELDS SEARCHED (Int.Cl.7) C12N C07K C12Q G01N A61K
INCOMPLETE SEARCH The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims. Claims searched completely : Claims searched incompletely : Claims not searched : Reason for the limitation of the search: Although claim 12, in as far as it pertains to in vivo use, is directed to a method of treatment of the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.			
Place of search		Date of completion of the search	Examiner
THE HAGUE		13 September 2001	Smalt, R
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			

EPO FORM 1503 03/82 (P04C07)

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 01 11 2637

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
P,X	WO 00 77173 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); RUBEN STEVEN M (US)) 21 December 2000 (2000-12-21) see seq.ID's 12, 53 and 117, and gene #2 -----	1,4-21, 23-26	
			TECHNICAL FIELDS SEARCHED (Int.Cl.7)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 01 11 2637

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

13-09-2001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0058473 A	05-10-2000	AU 3774500 A	16-10-2000
		WO 0058473 A2	05-10-2000

WO 0077173 A	21-12-2000	AU 5022200 A	05-12-2000
		AU 5174600 A	02-01-2001
		WO 0070076 A1	23-11-2000
		WO 0077173 A1	21-12-2000
